

A:Accession: A81919
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-447 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PID:CA884035.1; PID:g737947
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: ngrA; NNA0752
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquinone)
C:Keywords: NAD; oxidoreductase

alignment_scores:
Quality: 2211.00 Length: 447
Ratio: 5.059 Gaps: 0
Percent Similarity: 97.763 Percent Identity: 97.092

alignment_block:

US-09-303-518D-127 x A81919 ..

Align seg 1/1 to: A81919 from: 1 to: 447

```
1 ATGATTAATAAATAAGGCTCTAAAGCTGCCATCGCGGCGACAGCGGA 50
1 MetileLysileLysLysGlyLeuAsnLeuProileAlaGlyArgProG1 17
51 GCAAGTCATTTATGACGGGCGCGCTCATACCGAGTCGCGTTCGTTGGCG 100
17 uGlnValileTyrAspGlyProValIleThrGluValAlaLeuLeuGlyG 34
101 AAGAATATGCGGTPATCGGCGCTTNGATGAAAGTCAAGGAAGCGATGCC 150
34 LuGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50
151 GTCAAAAAGGCCAAGTCTGCTTTTGAAGACAAAAGNATCCGGGCGTGGT 200
51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67
201 GTTTACCGCCCGCTTCAGGCAAAATCGCGCATCCATCGCGCGGAAA 250
67 lPheThrAlaProValSerGlyLysIleAlaAlaIleHisArgGlyGluL 84
251 AGCGCGTACTTCAGTCTGCTGATGCTGCTGTAAGCGCAACGACGAAATC 300
84 ySArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100
301 GAGTTGAAACGCTACGCGCGCGAAGCGTTGGCAAACTTAAGCGCGGANGA 350
101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluG1 117
351 ANTNNNGCAATCTGATCCCAATCCGTTTGTGGACTCGCGTANCC 400
117 uValArgArgAsnLeuIleGlnSerGlyLeuTyrThrAlaLeuArgThrA 134
401 GTCCGTTTCAGAAAATCCCTGCGCTGATGCGGAGCGGCTTCGCCATCTC 450
134 rgProPheSerLysileProAlaValAspAlaGluProPheAlaIlePhe 150
451 GTCATGCGATGGACACCAATCCGCTNGCGCGACACCCCTGCTGTTCAT 500
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProValValIle 167
501 CAAGAAGCCGCGANGATTTTCAGACGANGTNTGCTGTATGAGCCGTT 550
167 eLysGluAlaAlaGluAspPheArgArgGlyLeuLeuValLeuSerArgL 184
551 TGACCGAGCGTAAATCCATGTTGTGAAGGAGCTGGCGAGAGCGTGGCG 600
184 euThrGluArgLysileHisValCysLysAlaAlaGlyAlaAspValPro 200
601 TCTGAAAATGCTGCCAATCGAATCGAATCGAATCGCGCGCGCATCC 650
201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyProHisPr 217
```

```
651 GGCGGTTTGTAGTGGCAGCGCATTCATTTCATTGAGCGGTCGGTCAA 700
217 oblaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAla 234
701 ACAAAACCGCTTTGGACCAATCAATATCAAGATGTAATTCGCCATCGGAGT 750
234 snLysThrValTrpThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250
751 TTGTTTGAACAGCGCGCTCTGAACACCGCGCTGATTCGTTGGTGGTGG 800
251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyG1 267
801 TTCTCAAGTCAACAAACACCGCGCTTTCGCTACCGTTTGGTGGTGGAG 850
267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284
851 TATCGCAAAATTAATCGCGCGCAATTTGGTTGACGACAGCAACCGCGTGA 900
284 alSerGlnIleThrAlaGlyGluLeuValAspAlaAsnArgValIle 300
901 TCCGTTTCGTTTCAACGCGCGGATTCACACAGCGCGCGCATTCATT 950
301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrFle 317
951 GGCAGCGCTACCAATCAAGATTCGCTTATCGAAGAGCGCGCGCAAG 1000
317 uGlyArgTyrHisAsnGlnIleSerValIleGluGlyArgSerLysG 334
1001 AGCTGTTCGGTGGTTCGCGCGCGAGCGGACAAATATCCATCAGCGT 1050
334 LuLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350
1051 AGCAACCTCGCGCATTCCTGAAAACAACTCTTCAAGTTTCACGACAGC 1100
351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheThrThrAl 367
1101 CGTCAACGTTGGCGCGCGCATTCGCTGCGGATTCGCTTACGACGCGC 1150
367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384
1151 TAATGCGCTAGACATCCCTGCGCTACCTGCTTTCGCGGATTCATTCGTC 1200
384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400
1201 GCGGATACGACGCGCGCAAGCATTCGCTTGGTTCGTTGGAATTCGACGA 1250
401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluG1 417
1251 AGACCTCGCTTTGTGACGCTTCGCTGCGCGCGCAAAATACGAATANGGCC 1300
417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434
1301 CGCTGTTCGTTAAGTTCGTTGGAACCTTCGAAAGGAGGCG 1341
434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447
```

seq_name: pir2:D81185

seq_documentation_block:

sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain NMB0569
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Mar-2001
C:Accession: D81185
R:Tetelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755
A:Accession: D81185
A:Status: preliminary

A;Molecule type: DNA
A;Residues: 1-447 <TIGR>

608 C A A C A A A C C C T T T G C C A C C A T C A A T T A T C A A G A T G T A A T T G C C A T C G G A 747

C;Genetics:
A;Gene: ngrA
C;Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquinol)
C;Keywords: oxidoreductase

[illegible]

us-09-303-518d-127.rpr

Mon Jul 1 09:25:28 2002

Quality:	1422.50	Length:	447
Ratio:	3.763	Gaps:	1
Percent Similarity:	84.564	Percent Identity:	61.074

alignment block:

US-09-303-518D-127 x S51015

Align seq 1/1 to: S51015 from: 1 to: 446

1 ATGATTAAATCAAAAAGGCTAAACCTGCCCATCGCGGCGACAGCCGA 50
1 MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyThrProSe 17
51 GCAAGTCATTATGACGGGCCGTCATTACACGAAGTCGCGTGTCTGGCG 100
17 rGlnValIleAsnAspGlyLysThrIleLysValAlaLeuLeuGly 84
101 AAGAATATCGCGTATGTCGCCCTCATGATGAAGTCAAGGAAGCGATGCC 150
34 LuGluTyrValGlyMetArgProThrMetHisValArgValGlyAspGlu 50
151 GTCAAAAAGGCGCAATGCTGTTTGAAGACAAAAGNATCCGGCGTGGT 200
51 ValLysLysAlaGlnValLeuPheGluAspLysLysAsnProGlyVal 67
201 GTTTACCGCCGCTTCAGGCAAAATCGCGCCATCATCGCGCGAAA 250
67 sPheThrAlaProAlaAlaGlyLysValIleGluValAsnArgGlyAla 84
251 AGCCGCTACTTCAGTCGGTGTGATTCCGCTTGAAGCAAGCAGCAATC 300
84 ysArgValLeuGlnSerValValIleGluValAlaGlyGluGlnVal 100
301 GAGTTCGAACGCTACGGGCCCGGAGCGTTGGCAAACTTAACGGCGGANGA 350
101 ThrPheAspLysPheGluAlaAlaGlnLeuSerGlyLeuAspArgGluVal 117
351 ANTNNNGCAATCTGATCAACATCCGTTTGTGGACTCGGCTGCGTANCC 400
117 lIleLysThrGlnLeuValAspSerGlyLeuThrPheAlaLeuArgThrA 134
401 GTCGCTTCAGCAAAATCCCTCGCTGCATGCCAGCGCTGCCCATCTTC 450
134 rgProPheSerLysValProAlaIleGluSerSerThrLysAlaIlePhe 150
451 GTCATATGGATGGACACCAATCCGCTNGCGGACAGACCCCTGTGTTGAT 500
151 ValThrAlaMetAspThrAsnProLeuAlaAlaLysProGluLeuIle 167
501 CAAGAAGCCGNCANGATTTCAGACGANGTGTGTTGATTGAGCCGCT 550
167 easnGluGlnGlnAlaPheIleAlaGlyLeuAspIleLeuSerAla 184
551 TGACGAGCGTAAATCCATGTGTGAAGCAGCTGGCGACAGCTGGCGG 600
184 euThrGluGlyLysValTyrValCysLys...SerGlyThrSerLeuPro 199
601 TCTGAANAATGTCACACATCGAACAACATGAATTCGGCGGCCCGCATCC 650
200 ArgSerSerGlnSerAsnValGluGluHisValPheAspGlyProHisPr 216
651 GCCCGTTTGTAGTCGACGCGACATTCATTTCATTGACCGCGTGGTAA 700
216 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyrProValAsnAla 233
701 ACAAAAACGTTTGGACCATCAATTCATCAAGATGTAATGCCATCGGACGT 750
233 LuAsnValAlaTrpSerIleAsnTyrGlnAspValIleAlaPheGlyLys 249
751 TTGTTTGAACAGGCCGCTCTGAACACCGACGCGGTGATTGCTTTGGGTG 800
800 LeuPheLeuThrGluLeuTyrThrAspArgValValSerLeuAlaIle 866
866 LeuPheLeuThrGluLeuTyrThrAspArgValValSerLeuAlaIle 866

Percent Similarity: 83.893 Percent Identity: 60.403

alignment_block:

US-09-303-518D-127 x G82094

Align seg 1/1 to: G82094 from: 1 to: 462

```
1 ATGATTAATAATCAAAAGGCTCTAAACCTGCCCATCGCGGCGAGACCGGA 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
17 MetilethrelylsylsGlyLeuaspLeuProIleAlaGlyThrProse 33
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 GCAAGTCATTATGACGGCGCGCTCATTAACCAAGTCGCGTTCCTGGCG 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
33 rGlnValIleSerAspGlyLysAlaIleLysLysValAlaLeuLeuGlyG 50
101 AGAATATATCGCGGTATGCGCCCTGATGAAAGTCAAGGAAGCGCATGCC 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
50 luGluTyValGlyMetArgProThrMethIsvAlargValGlyAspGlu 66
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 GTCAAAAAGGCCAAGTGTCTGTTGAAGACAAAAAGNATCCGGCGTGT 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
67 ValLysLysAlaGlnIleLeuPheGluaspLysLysAsnProGlyVally 83
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
201 GTTTACCGCGCGNGTTCAGGCAAAATCCGCCCATCCATCGCGCGGAAA 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
83 sPheThrSerProValSerGlyLysValValGluIleAsnArgGlyAlaL 100
251 AGCGGTACTTCAGTCGCGTGTGATGCGGTGGAAGCAACGACGCAAAATC 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
100 ysArgValLeuGlnSerValValIleGluValAlaGlyaspAspGlnVal 116
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
301 GAGTTCGACGCTACGCGCGCGAAGCGTTGGCAACTTAAGCGCGGANGA 350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
117 ThrPheAspLysPheGluAlaAsnGlnLeuAlaSerLeuAsnArgAspAl 133
351 ANTNNNGNCAATCGATCCATCGGTTTGTGACTCGCGTGGGTANCC 400
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
133 arlelysthrGlnLeuValGluSerGlyLeuTrpThrAlaPheArgThrA 150
401 GTCCGTTCCAGCAAAATCCCTGCCGTCGATCGCGCGCGTTCGCCATTC 450
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
150 rgProPheSerLysValProAlaIleaspSerThrSerGluAlaIlePhe 166
451 GTCATCGGATGGACACCAATCCGTCNGCGCGCAGACCGCTGTGTGTGAT 500
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
167 ValThrAlaMetAspThrAsnProLeuAlaAlaGluProThrValValII 183
501 CAAAGAAGCCGCGANGATTTTCAGACGANGTNTCTGTGTTATGACCGTT 550
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
183 eAsnGluGlnSerGluAlaPheValAlaGlyLeuaspValLeuSerAlaL 200
551 TGACCGAGCGTAAATCATGTGTGTGAAGCAGCTGCGCGCAGAGTGGCG 600
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
200 euThrThrGlyLysValTyValCysLys...LysGlyThrSerLeuPro 215
601 TCTCAAAATGTCGCAACATCGAACAACATCAATTCGCGCGCGCGCATCC 650
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
216 ArgSerGlnGlnProAsnValGluLuhIsvAlpheaSpGlyProHisPr 232
651 GGCGGTTTGTAGTGGCAGCACATTCATTTCATGAGCGGTGGGTGCAA 700
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
232 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyProValSerAlaA 249
701 ACAAACCGTTTGCACCATCAATTAACATGATTAATTCGCAATCGGACGT 750
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
249 sPHisValAlaIleAsnTyrglnAspValIleAlaValGlyGln 265
751 TTGTTTCAACAGCGCGCTCTGAACACGCGGTGATTCCTTGGGTGG 800
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
266 LeuPheLeuThrGlyGluLeuTyThrGlnArgValValSerLeuAlaGl 282
801 TTCTCAAGTCAAAACACGCGCTTGTGCGTACCGTTTGGGTGCGAAAG 850
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```
282 yProValValAsnLysProArgLeuValArgThrValMetGlyAlaSerL 299
851 TATCGCAAATTAATCGGGCGAATTTGACGACGACCAACCGCGTGTAT 900
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
299 euGluGlnLeuValAspSerGluIleMetProGlyGluValArgIlelle 315
901 TCCGGTTCGGTATTGACGGCGCGGATTAACAAGCGCGCACGATTATTT 950
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
316 SerGlySerValLeuSerGlyThrLysAlaThrGlyProHisAlaTyLe 332
951 GGGACGCTACCAATCAGATTTCCGTTATCAAGAAGCGCGCACGAAG 1000
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
332 uGlyArgTyHisLeuGlnSerValLeuArgGluGlyArgAspLysG 349
1001 AGCTGTTGCGTGGTTCGCGCGCAGCGGACAAATACTCCTCACCGCT 1050
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
349 luLeuPheGlyTrpAlaMetProGlyLysAsnLysPheSerValThrArg 365
1051 ACGACCTCGCCCATTTCTCTGAAACAAACTCTTCAAGTTTCACGACAG 1100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
366 SerPheLeuGlyHisLeuPheLysGlyGlnValTyArgMetThrThr 382
1101 CGTCAACGGTGGCGACCGCATGTCGCGATTTGCTGCTGCTGCTGCTG 1150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
382 rThrAsnGlySerAspArgSerMetValProIleGlyAsnTyrgLulysv 399
1151 TAATGCGCGCTAGACATCCTCCTACCTGCTTTTGGCGCGGATTAATCG 1200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
399 alMetProLeuAspMetGluProThrLeuLeuLeuArgAspLeuCysAla 415
1201 GCGGATACCGACGCGCAGCATTTGGGTGCTTGGATTTGGACGACGA 1250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
416 GlyAspSerAspSerAlaValArgLeuGlyAlaLeuGluLeuAspGlu 432
1251 AGACCTCGCTTTGTGAGCTTCGTCGCCGCGCAATACGAATANGGCC 1300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
432 uAspLeuAlaLeuCysThrPheValCysProGlyLysTyrgLulysv 449
1301 CGCTGTTGCTAAGTTCGCGAAACCCNTTGGAGAAAGGAGGC 1341
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
449 lnLeuLeuArgGluCysLeuAspLysIleGluLysGluGly 462
```

seq_name: pir2:H83272

seq_documentation_block:

sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) nqr1 chain PA2999 [C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 16-Feb-2001
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: H83272
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-445 <STO>
A:Cross-references: GB:AE004724; GB:AF004091; NID:g9949083; PIDN:AAG06387.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: nqrA; PA2999
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino
C:Keywords: Oxidoreductase

alignment_scores:

Quality: 1298.00 Length: 447
Ratio: 3.576 Gaps: 1
Percent Similarity: 81.208 Percent Identity: 57.271

alignment_block:

```
US-09-303-518D-127 x H83272
Align seg 1/1 to: H83272 from: 1 to: 445

1 ATGATTAAATCAAAAGAGTCTAAACCTGCCATCGCGGCGACACCGGA 50
1 Metilelylelyysargglyleuaspleuproilleseriglyalaprog 17
51 GCAAGTCATTATGACGGGCGCGCTTATACCGAAGTCGCGTGTGGCG 100
17 uGlnArgIleGluAlaAlaArgProValArgSerValAlaLeuIleGlyP 34
101 AGAATATCGCGGTATCGCCCTCTGATGAACTCAAGGACCGCATGCC 150
34 heAspTyrHisGlyMetLysProThrMetAlaValGlnValGlyAspArg 50
151 GTCAAAAGAGCGCAAGTCGTGTTGAAGACAAAGNATCCGGCGCTGT 200
51 ValLysLeuGlyGlnValLeuPheThrAspLysLysAsnProSerValSe 67
201 GTTTACCGCGCCNGTTTCAGGAAATCGCCGCCATCCATCGCGGGGAA 250
67 rTyrThrAlaProGlyAlaGlyValValSerAlaIleHisArgGlyGlu 84
251 AGCGCGTACTTCAGTCGTCGATTCGCGTGTGAAGGCAACGACGAAATC 300
84 ysArgValLeuGlnSerValValIleAspLeuAspGlyAspGlnLeu 100
301 GAGTTCGAACGCTACGCGCCCGAGCGTGTGGCAAACTTAAGCGCGGANGA 350
101 GluPheAlaArgTyrProAlaAspLysLeuAlaThrLeuSerAlaGluG 117
351 ANTNGNNGCAATCGATCCAAATCCGCTGTGTGGACTCGCGTANCC 400
117 nValArgAspAsnLeuLeuGlnSerGlyLeuTrpThrAlaLeuArgThrA 134
401 GTCCGTTTCAGAAATCCCTGCCGTCGATCGGAGCGCTTCGCCATCTTC 450
134 rgpPropheSerLysValProAspProgluSerSerProSerIlePhe 150
451 GTCAATCGGATGGACCAATCCGCTNGCGGACACCTGTGTGTGTGAT 500
151 ValThrAlaIleAspThrGlnProLeuAlaAlaAspProGlnValValII 167
501 CAAGAAGCCGNCANGATTTTCACAGCANGTCTGCTGTTAGCGCGTT 550
167 eAlaGluGlnGlyGluAlaPheGlnAlaGlyLeuThrValLeuGlyArgL 184
551 TGACCGAGCGTAAATCCATGTGTGTAAGCGAGCTGGCGCAGACGTCGC 600
184 euAla.....ArgValPheLeuLysLysAlaGluGlyValSerLeuPro 198
601 TCTGAAATGTCGCAACATCGAACACATGAATTCGGCGCGCCGCATCC 650
199 GlyGluAlaLeuSerGlyValThrAlaGlnAlaPheSerGlyProHisPr 215
651 GCCTGTTTGTGAGCGCAGCATTCATTTCATTGAGCGCGCTCGGTGCAA 700
215 oAlaGlyLeuProGlyThrHisIleHisPheLeuAspProValGlyAlag 232
701 ACAAAACCGTTTGGACCATCAATTAATCAAGATGAATTCATCGCGAGCT 750
232 lYlSerValTrpAsnLeuAsnTyrGlnAspValIleAlaIleGlyLys 248
751 TTGTTTGAACAGCGCTCTGAACAGCGCGCTGATTGCTTGGGTGG 800
249 LeuPheThrThrGlyGlnLeuTrpThrGluArgValIleAlaLeuAlag 265
801 TTCTCAAGTCAACAAACACGCTCTTCGTCACCGCTTTTGGGTGCGAAAG 850
265 yProValValGluLysProArgLeuValArgThrArgLeuGlyAlaAsnL 282
851 TATCGCAATTTACTCGGGCGAATGTTGGTTGACGCGACACACCGCGTATT 900
```

```
.....:|||||:|||||:.....:|||||:|||||:|||||:|||||
282 euAspGluLeuAlaAlaGlyGlnLeuGlnProGlyAsnAsnArgLeuIle 298
901 TCCGGTTCGGTATTGAACGGCGGCGATTACACAGGCGCGCACGATTTATTT 950
299 SerGlySerValLeuGlyGlyArgThrAlaHisGlyAlaTyrAlaTyrLe 315
951 GGGACGCTACCACCAATCAGATTTCCTGTTATCGAAGAAGCGCGCAGCAAAG 1000
315 uGlyArgTyrHisLeuGlnLeuSerCysLeuLysGluGlyAspGlnArg 332
1001 AGCTGTTTCGGTGGTGGTGGCGCGCGGACAAATCTCAAGTTTCACGACGC 1050
332 LuPheLeuHisTyrLeuArgAlaGlyValGluLysHisSerLeuLeuAsn 348
1051 AGGACCTTCGGCCATTTCCTGAAAAACAACTCTCAAGTTTCACGACGC 1100
349 ValPheValSerArgLeuLeuGlyGlyLysArgPheAlaPheThrThrSe 365
1101 CGTCAACGCTGGCGACCGCGCATGGTGGTGGTGGTGGTGGTGGTGGTGG 1150
365 rThrAsnGlySerProArgAlaMetValProValGlyAsnTyrGluAlaV 382
1151 TAATCGCGCTAGACATCTCTGCTACCTCCCTGCTTTCGCGGATTAATCGTC 1200
382 alMetProLeuAspIleLeuProThrGlnLeuLeuArgTyrLeuIleVal 398
1201 GCGCATACCGACGCGCGCAGCATGGTGGTGGTGGTGGTGGTGGTGGTGG 1250
399 GlyAspThrGluMetAlaGlnLysLeuGlyAlaLeuGluLeuAspGluG 415
1251 AGACCTCGCTTTGTGCGAGCTTGTCTCCCGCGGCAAAATACGAATANGGCC 1300
415 uAspLeuAlaLeuCysSerTyrValCysAlaGlyLysTyrGluTyrGlyP 432
1301 CCCTGTTGCTAAGTGTCTGGAACACCTNTGAGAAGGAAGGC 1341
432 roIleLeuArgAspAsnLeuAlaAlaArgIleGluGlnGluGly 445
seq_name: pir2:D86583

seq_documentation_block:
ubiquinone oxidoreductase, alpha [imported] - Chlamydothilla pneumoniae (strain J138)
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: D86583
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: D86583
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-467 <STO>
A:Cross-references: GB:BA000008; NID:98979115; PIDN:BAA98950.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
C:Gene: nqrA
C:Supertfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino

alignment_scores:
Quality: 447.50 Length: 468
Ratio: 1.565 Gaps: 15
Percent Similarity: 61.111 Percent Identity: 29.487

alignment_block:
US-09-303-518D-127 x D86583
Align seg 1/1 to: D86583 from: 1 to: 467
4 ATTAATAATCAAAAGAGTCTAAACCTGCCATCGCGGCGACACCGGACGA 53
||| :.....:|||||:|||||:|||||:|||||:|||||:|||||
```



```
3 IleThrValAsnArgGlyLeuAspLeuSerLeuGlnGlySerProLysG1 19
54 AGTCATTATTATGACGGCCGCGTATACCAAGCGGCGGCGGAG 103
19 uSerGlyPheTyrAsnLysIleAspProGluPheValSerIle.....33
104 AATATGCCGGTATGGCCCG.....TNGATGAAGATC 135
34 .....AspLeuArgProPheGlnProLeuSerLeuLysLeuVal 47
136 AAGGAGGCGATCGCGTCAAAAAGCGCAAGTCGTTTGAAGACAAAA 185
48 GluGlnGlyAspAlaValCysSerGlyAlaProIleAlaGluTyrLysH1 64
186 GNATCGGCGCGTGGTGTATACCGCCGCTTTCAGSCAAATCGCCGCA 235
64 sPheProAsnThrTyrIleThrSerHisValSerGlyValValThrAlaI 81
236 TCATCGCGCGGAAAGCGCGTACTTCAGTCGGTGGTATT...GCCGTT 282
81 leArgArgGlyAsnLysArgSerLeuLeuAspValIleIleLysLysThr 97
283 GAAGCAACAGGAGCAATCGATCCAGCTACCGCCCGCGGAGCGTTCGC 332
98 ProGlyProThrSerThrGluTyr.....ThrTyrAspLeuG1 110
333 AAACCTAAGCGCGGANGAANTNNGCAATCTGATCAATCCGCTTGT 382
110 nThrLeuSerArgSerAspLeuSerGluIlePheLysGluAsnGlyLeuP 127
383 GGATCGCGCTCGGTANCGTCCGTTACGAAATCCCTCGCGTCGATGCC 432
127 heAlaLeuIleLysGlnArgProPheAsp...IleProAlaIleProThr 142
433 GAG...CGTTTCGCGATCTTCGTCGATCGATGACACCAATCCGTCNGC 479
143 GlnThrProArgAspValPheIleAsnLeuAlaAspAsnArgProPheTh 159
480 GCGAGACCGCTGTGGTGTGATC.....AAAGAAGCGG 511
159 rProSerProGluLysHisLeuAlaLeuPheSerSerArgGluGluGlyP 176
512 NCGANGATTTCAGACGANGTNGTGGTATTGACCGCTTTCAGCGAGCGT 561
176 heTyrValPheValValGlyValArgAlaIleAlaLysLeuPheGlyLeu 192
562 AAAATCCATGTGTGAAGCGAGCTGGCGCAGACGTCGCGTCAAAATGC 611
193 ArgProHisIleValPheArgAspArgLeuThrLeuProThrGlnGluLe 209
612 TGCCCAACATC...GAACACATGAATTCGGCGCCCGCCGATCCGCGCGTT 658
209 uLysThrIleAlaHisLeuHisThrValSerGlyProPheProSerGlys 226
659 TGAGTGGCAGCAGCATTCATTTCATTCAGCGCGTGGTGCACAAACA... 705
226 erProSerIleHisIleHisSerValAlaProIleThrAsnGluLysGlu 242
706 ACCGTTTGGACCATCAATTAAGATGATTAATTCATCCATCGAGGTGTTT 755
243 ValValPheThrLeuSerPheGlnAspValLeuThrIleGlyHisLeuPh 259
756 TGCAACAGCGCGTCTGAACACCGCGCGCGTATGCTTGGTGGGTCTC 805
259 eLeuLysGlyArgIleLeuHisGluGlnValThrAlaLeuAlaGlyThrA 276
806 AAGTCAACAAACCA.....CGCGCTCTTGGCTACCGGTTTGGTGGCGAAA 849
276 laLeuLysSerSerLeuArgArgTyrValIleThrThrLysGlyAlaSer 292
850 GTATCGCAA...ATTACTCGCGCGGATGTTGGTTGACGACAGAACCGCGT 896
293 PheSerSerLeuIleAsnLeuAsnAspIleSerAspAsnAspThr...Le 308
```

```
897 GATTTCGGTTTCGTTATTGAACGGCGGATACACAAGCGCGCGCAGCAT. 945
308 uIleSerGlyAspProLeuThrGlyArgLeuCysLysLysGluGluGluP 325
946 ..TATTGGGACGCTACCAATCAGATTTCGTTATCGAAGAGGCGCG 993
325 roPheLeuGlyPheArgAspHisSerIleSerValLeuHisAsnProThr 341
994 AGCAAAAGAGCTGTTTCGGCTGGCGCCGCGCAGCCGCGCAAAATCTCCAT 1043
342 LysArgGluLeuPheSerPheLeuArgIleGlyPheAsnLysProThrPh 358
1044 CACCGCTACGACCGCTCGGCCATTTCCTTGAAAAACAAACTCTTCAAGTTCA 1093
358 eThrLysThrTyrLeuSerGlyPhePheLysLysArg...ThrTyrT 374
1094 CG.....ACAGCGGTCAACGGTGGCGACCGCGCGCATGGTGGCGATT 1134
374 hrAsnProAspThrAsnLeuHisGlyGluThrArgProIleIleAspThr 390
1135 GGTACTTACGAGCGGTAATCGCGCTAGACATCTGCTACCCCTGCTTTT 1184
391 AspIleTyrAspLysValMetProMetArgIleProValValProLeuI 407
1185 GCGCGATTATTCGTCGGCGATACCGACGCGCGCAAGCATTTGGTGTGT 1234
407 eLysAlaValIleThrLysAsnPheAspLeuAlaAsnGluLeuGlyPheL 424
1235 TGGAAATTGGACGAAGACCTGCTTGTGCGAGCTTCGTCGCGCGCGG 1284
424 euGluValCysGlyGluAspPheAlaLeuProThrLeuIleAspProSer 440
1285 AAATACAGATANGCGCGCTGTTGCGTAAGTCTCGTGAACACCTTTGAGAA 1334
441 LysThrGluMetLeuThrIleValLysGluSerLeuIleGluTyrAlaL 457
1335 GGAA 1338
457 sGlu 458
```

seq_name: pir2:E72040

seq_documentation_block:

probable sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 16-Feb-2001
C:Accession: E72040; G81623
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: E72040
A:Molecule type: DNA
A:Residues: 1-467 <AR>
A:Cross-references: GB:AE001656; GB:AE001363; NID:g4377047; PIDN:RAD18882.1; PID:g437
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: G81623
A:Molecule type: DNA
A:Residues: 1-467 <REA>
A:Cross-references: GB:AE002164; GB:AE002161; NID:g7188939; PIDN:AAF37899.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: nqrA; CP0002
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino
C:Keywords: oxidoreductase

756 TGCAACAGCGCGTCTGAACACCAGGCGCGTGATGTGTTGGGPGTTC 805
| | | | | : : : | | | : : : | | | : : : | | | : : :
259 eLeuLysGlyArgIleLeuHisGLuInValThrAlaLeuAlaGlyThrA 276
| | | | | : : : | | | : : : | | | : : : | | | : : :
806 AAGTCAACAACCA.....CGCCTCTTGCCTACCGTTTGGTGGCGAA 849
: : : : : : : : | | : : : | | | : : : | | | : : :
276 laLeuLysSerSerLeuArgArgtyrValIleThrThrLysGlyAlaSer 292
| | | | | : : : | | : : : | | | : : : | | | : : :
850 GTATCGCAA...ATTACTCGGCGGAATTGGTTGCAGCACACACGCCGT 896
| | | : : : | | | : : : : : : : : | | | : : : : : :
293 PheSerSerLeuIleAsnLeuAsnAspIleSerAspAsnAspThr...Le 308
| | | : : : | | | : : : : : : : : | | | : : : : : :
897 GATTTCGCGTTTCGATTATTAACGCGCGGATTTACACAGGCGCGACGAT.. 945
: | | | | : : : : : | | | : : : : : : : : : : : : : : :
308 uileSerGlyAspProLeuthrGlyArgLeuCysLysLysGLuGLuGlup 325
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
946 ..TATWTGGGACGCTACCAcAATCAGATTTCCGTTATTCGAAGAAGCCGC 993
: | | | | : : : : : | | | : : : : : : : : : : : : : : :
325 roPheLeuGlyPheArgAspHisSerIleSerValLeuHisAsnProThr 341
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
994 AGCAAAAGAGCTGTTCGCTGGGTGGCGCGAGCGCGACAAAATACTCCAT 1043
: : : : : | | | | | : : : : : : : : : : : : : : : :
342 LysargGLuLeuPheSerPheLeuArgILeGlyPheAsnLysProthrPh 358
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
1044 CACGCGTACGACCCTCGGCCATTTTCCTGA AAAACA AA ACTCTTCAAGTWCA 1093
: | | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
358 eThrLysThrtyrLeuSerGlyPhePheLysLysLysArg...ThrrTyrt 374
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
1094 CG.....ACAGCCGTCAACCGTGGCGNCCGCGCATGGTCCGATT 1134
| | | : : : : : | | | : : : : : | | | : : : : :
374 hrAsnProAspThrAsnLeuHisGlyGLuThrArgProIleIleAspThr 390
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
1135 GGTACTTACGAGCGGTAAATCCGCTAGACATCTCGCTACCCCTGCTTTT 1184
| | | : : : : : | | | : : : : : | | | : : : : :
391 AspIleItyrAspLysValMetPrometaArgIleProValValProLeuil 407
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
1185 GCGCGATTAAATCGTCGCGCATACCGACAGCGCGCGAACGATTGGTGGTCT 1234
: | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
407 elysAlaValIleThrLysAsnPheAspLeuAlaAsnGLuLeuGlyPheL 424
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
1235 TGGAAATTCGACAAGAACACCTCGCTTGTGTCAGCTTCGTCTGCCCGGGC 1284
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
424 euGLuValIcysGlyGLuaspPheAlaLeuProthrLeuIleAspProSer 440
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
1285 AAATACGAAATANGGCCCTCTTGCCTAAGGTGTGGAAACCNNTTGAA 1334
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
441 LysThrGLuMetLeuThrIleValLysGLuSerLeuIleGLuTyraLaLy 457
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
1335 GGAA 1338
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
457 sglu 458

seq_documentation_block:
probable sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 16-Feb-2001
C:Accession: C81751
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: C81751
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-465 <YTT>
A:Cross-references: GB:AE002269; GB:AE002160; NID:g7190041; PIDN:AAF38896.1; PID:g719
A:Experimental source: strain Nig9 (MoPn)
C:Genetics:
A:Gene: TC0002

C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquinone)
C:Keywords: oxidoreductase

alignment_scores:
Quality: 441.00 Length: 464
Ratio: 1.521 Gaps: 14
Percent Similarity: 62.500 Percent Identity: 28.448
alignment_block:
US-09-303-518d-127 x C81751
Align seg 1/1 to: C81751 from: 1 to: 465

```
4 ATTAATCAAAAAAGGCTCTAAACCTGCCATCGCGGCGACGACGCGGAGCA 53
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
3 IleileValSerArgGlyLeuAspLeuSerLeuLysGlyAlaProLysGI 19
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
54 AGTCATTATGACGGG.....CCGCTCATTCAGCAATCGCGTGC 94
: : : : : : : : : : : : : : : : : : : : : : : : : : :
19 userGlyPheCysGlyLysValAspProAlaPheValSerValAspLeu. 35
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
95 TTGGCGAAGAATATGCGCGTATGCGCCCTNGATGAAAGTCAAGGAGGC 144
: : : : : : : : : : : : : : : : : : : : : : : : : : :
36 .....ArgProPheAlaProLeuProLeuGlyValLysValSerProGI 50
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
145 GATCGCGTCAAAAAAGGCGCAAGTCTGTTGAAGACAAAGNATCGGG 194
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
51 AspGlnIleThrAlaGlySerProLeuAlaGluTyrLysSerPheProGI 67
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
195 CGTGGTGTATACCGCGCGCTTCAGCAAAATCGCGCCATCATCGCG 244
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
67 yValPheIleThrSerValAspGlyGluValIleCulIleArgG 84
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
245 GCGAAAAAGCGGTACTTCAGTGTGCTGATGCGGTGAA...GGCAAC 291
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
84 lySerLysArgAlaLeuLeuAspIleValIleLysLysLysProGlyVal 100
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
292 GACGAATCGATTGCAAGCGTACGCGCCGCAAGCGTTGGCAACTTAAG 341
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
101 SerGlnThrLysPheSer.....TyrAspLeuHisAlaLeuSe 113
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
342 CGCGGANGAANTNNGNCAATCTGATCCATCGGTTTGTGGACTCGC 391
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
113 rGlnLysGluLeuLeuGluValPheLysLysGluGlyLeuPheThrLeuP 130
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
392 TGCATNCCGTCGCTTCAGCAAAATCCCTCCGTCGATGCGGAG...CCG 438
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
130 heLysGlnArgProPheAsn...IleProAlaLeuProThrGlnSerPro 145
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
439 TTCGCCATCTTCGTCATGCGATGGACACCAATCCGCTN.....CGGGC 482
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
146 ArgAspValPheIleAsnLeuAlaAspAsnArgProPheThrProSerVa 162
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
483 AGACCTGTGTGTGTGATCAAGAAAGCGCGCGANGAT.....T 520
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
162 lGluLysHisLeuSerLysPheSerSerLysGluAspGlyTyrTyrIleP 179
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
521 TCAGACGANTGTGCTGTTATGAGCGGTTGACCGCGCTAAATCCAT 570
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
179 heValValGlyValGlnAlaIleAlaLysLeuPheGlyLeuLysProHis 195
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
571 GTGTGTAAGGCACCTGCGCGACAGCTGCGCTCTGAAATGCTCCCAACAT 620
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
196 IleValSerThrAspArgLeuSerLeuProThrGlnAspLeuSerVa 212
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
621 C...GAACACATGAATTCGGCGCGCGCATCGCGCGGTTTGAAGTGCA 667
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
212 lAlaHisLeuHisThrIleAlaGlyProThrProSerGlySerProSerT 229
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
668 CGCACATTCAATTCATGACCGGTCGCGTGCAGCAACAAA...ACCGTTGG 714
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
229 hrHisIleHisIleAlaArgIleArgAsnAspArgAspIleValPhe 245
```

```
715 ACATCAATATCAAGATGTAATTCATCGACGCGGTTGTTGCAACAGG 764
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
246 ThrIleSerPheGlnGluValLeuSerIleGlyHisLeuPheLeuLysGI 262
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
765 CCGTCTGAACACACGCGCGTGTGATTGCTTTGGGTGCTTCTCAAGTC... 810
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
262 yPhePheLeuGlyGlnGlnValAlaLeuAlaGlySerAlaLeuProp 279
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
811 ..AACAAACACACGCTCTTTCGTACCGTTCGTTGGTGGTAAAGTATCGCAA 858
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
279 roSerGlnArgLysTyrLeuIleThrAlaLysGlyAlaSerPheLysAsp 295
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
859 ATTACTCGCGGCGAATGTTGTGACGCA...GACACCGCGTGTATTCGCG 905
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
296 LeuLeuProGlnGluIlePheSerSerAsnAspValSerLeuIleSerGI 312
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
906 TTCGGTATTGACGCGCGGATTT...ACACAAGCGCGCACGATTATTGCG 952
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
312 YAspProLeuThrGlyArgLeuCysAsnLysGluGluAsnProCysLeuG 329
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
953 GAGCGTACCACAAATCAGATTTTCGTTATCGAAGCGCGCGCAAGAGAG 1002
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
329 lyMetArgAspHisThrIleThrIleLeuProAsnProLysThrArgGlu 345
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1003 CTGTTTCGGTGGTTCGCGCGCGACCGGACAAATCTCATCGCGGTAC 1052
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
346 MetPheSerPheLeuArgLeuGlyTyrAsnLysLeuThrValThrArgth 362
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1053 GACCTTCGCGCATTTCTCGAAAAACAAA.....CTCTTCAAGTTCAACA 1096
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
362 rTyrLeuSerGlyPhePheLysArgLysArgValPheMetAspMetAsnT 379
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1097 CAGCGTCAACGGTGGCGACCGCGCATGTCGCGATTGGTACTTACGAG 1146
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
379 hrAsnLeuHisGlyGluLysArgProIleIleAspSerGluIleTyrGlu 395
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1147 CGCGTAATCGCGTAGACATCTGCTACCTGCTTTTTCGCGGATTAAT 1196
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
396 LysValSerAlaIleAlaValProValAlaProLeuIleLysAlaLeuGI 412
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1197 CGTGGCGATACCGACGCGCAAGCATTTGGTTGCTTGGAAATGGACG 1246
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
412 uThrGlnAsnPheGluGluAlaCysArgLeuGlyLeuLeuGluValSerP 429
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1247 AAGAGACCTCGCTTGTGACACTTCGCTGCGCGCGCAAAATACCAATAN 1296
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
429 roGluAspPheAlaLeuProThrPheIleAspProSerLysThrGluMet 445
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1297 GCGCGCTGTTTCGTAAGTGTGTAACACCTTTGAGAAGGAA 1338
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
446 PheAlaIleValLysGluAlaLeuIleArgTyrAlaLysGlu 459
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

seq_name: pir2:F71489

seq_documentation_block:

sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain [simila
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 02-Mar-2001
C:Accession: F71489
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: F71489
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-465 <RNR>
A:Cross-references: GB:AE001334; GB:AE001273; NID:g3329078; PIDN:AAC68238.1; PID:g332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: nraA
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino

100

R.Kasahara, S.; Yamada, H.; Mio, T.; Shiratori, Y.; Miyamoto, C.; Yabe, T.; Nakajima, T.
J. Bacteriol. 176, 1498-1499, 1994
A:Title: Cloning of the *Saccharomyces cerevisiae* gene whose overexpression overcomes the
A:Reference number: A53382; MUID:94156857

A:Accession: A53382
A:Molecule type: DNA
A:Residues: 1-581, 'A', 583-593, 'A', 595-1802 <KAS>
A:Cross-references: EMBL:S69101; NID:9545659; PIDN:AAB30051.1; PID:9545660
A:Experimental source: YNN295
A:Note: sequence extracted from NCBI backbone (NCBIN:144410, NCBIIP:144411)
C:Genetics:

A:Gene: SGD:HKR1
A:Cross-references: SGD:S0002828; MIPS:YDR420w
A:Map position: 4R

C:Keywords: calcium binding; glycoprotein; transmembrane protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1802/Product: HKR1 protein #status predicted <NAT>
F:1483-1508/Domain: transmembrane #status predicted <TMW>
F:1645-1656/Domain: calcium binding #status predicted <CA1>

alignment_scores:
Quality: 133.50 Length: 445
Ratio: 0.588 Gaps: 20
Percent Similarity: 51.011 Percent Identity: 22.921

alignment_block:
US-09-303-518d-127/rev x S69703

Align seg 1/1 to: S69703 from: 1 to: 1802

```
1337 TCCTCTCAANGTTTCCACACCTTACGCAACAGCGGGCCNATTCGTA 1288
|||||
390 SerAspThrValValSerThrAsnThrValProTyrSerSe 406
1287 TTTCGCCGGCAGACGAGGTGCACAAAGCGAGGTCTTCTCGTCAAT 1238
|||||
406 rVal...HisSerThrPheValHis...AlaThrSerSerThrTyrI 421
1237 CCAAGCAACCAATGCTTGGCGCTGTCGATCGCGAGCATTAATCG 1188
|||||
421 LeSerSerSerLeuTyrSerSerProSerLeuSerAlaSerValSer 437
1187 CGCAAAAGCAGGTGAGGATGTAGCGGATTTACGCGCTCGTAAGT 1138
|||||
438 HisPheGlyValAlaProPheProSerAlaTyrIleSerPheSerVa 454
1137 ACCAATC.....GGCACCATGGCGGTGCGCCACCGTTGACGCGTG 1097
|||||
454 lProValAlaValSerSerThrTyrThrSerSerProSerAlaSerVal 471
1096 TCGTGAACCTTGAAGAGTTGTTTTCAGAAATGGCGAGGTCGTACGC 1047
|||||
471 alVal..... 472
1046 GTGATGGAGTATTGTCGGCTGGCGGCACCCAGCGCAACAGCTCTTT 997
|||||
473 .....ProSerAlaTyrAlaSerSerProSerValProVa 484
996 GTCGGCGCTTTCGATAACGGAATCTGATTGTTGGTACGTCCTCAAT 947
|||||
484 lAlaValSerSerThrTyrThrSerSer..... 493
946 AATCGTGGCGCTGTGTGTAATC...GGCGGTTCAATACGAAACCGGAA 900
|||||
494 ..ProSerAlaProAlaAlaSerSerThrTyrThrSerSerProSer 509
899 ATC.....ACCGGTTGTCTCGCTCAACCAATTCGCCCGC 865
|||||
510 AlaProValAlaValSerSerThrTyrThrSerSerProSerAlaProAl 526
864 AGTAATTTGCCATCTTTC.....GCACCCAAACGGTAC 830
|||||
```

seq_name: pir2:AE0692

seq_documentation_block:

probable NADH reducing dehydrogenase STY1665 [imported] - Salmonella enterica subsp.
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001

```
526 aAlaIleSerSerThrTyrThrSerSerProSerAlaProValAlaVal. 542
829 CCAAGAGCGGTGTTTGTGACTTGAACAACA.....CCCAAGCAATC 786
543 .....SerSerThrTyrThrSerSerProSerAlaProAlaAlaIle 556
785 ACGCGCTCGGTGTTTCAGACGGCTGTTGCAACAACAGTCCGATGGCAAT 736
557 SerSerThrTyrThrSerSerProSerAla.....ProValAlaVa 570
735 TACATCTTGATAATTGATGTCGAACGGTTCCTTTGTCACCGACCGCT 686
570 lSerSer.....ThrTyrThrSerSerProSerAla 581
685 CAATGAATGAATGTGCGTGCCTCAACACGGCGGATGCGGCGCGCG 636
581 ro.....ValAlaIleSerSerThrTyrThrSerSerPro 592
635 AATTCAATGTTTCGATGTTGGCAGCATTTTCAGACGGCAGCTCTGCGCC 586
593 SerValProValAlaValSerSerThrTyrThrSerSerProSerAlaPr 609
585 AGCTGCCCTTACACACATGGATTTTACGCTCGTCAAA.....CGGC 545
609 oAlaAlaIleSerSerThrTyrThrSerSerProSerAlaProValAlaV 626
544 TCAATACCAGCANACNTCGTCTGAATCNTCGNCGGCTTCTTTGATCACA 495
626 alSerSerThrTyrThrSerSerProSerAlaProAlaAlaIleSerSer 642
494 ACCACAGGTCGCGCGNAGCGGATGTTGTCCTCCATCGCATTCGACGAGAT 445
643 ThrTyrThrSerSerProSerValProValAlaValSerSerThrTyrTh 659
444 GCGCAACGGCTCGCATCGACGGCAGGATTTTCTGAACGGCAGCGGNTAC 395
659 rSerSerProSerAlaProAlaAla.....IleSerSerThrTyrT 673
394 GCGCGCAGTCACAAACCGGATTCGATCAGATTCGNCNNTTCNTCWC 345
673 hrSerSerProSerValProValAlaValSerSerThrTyrThrSerSer 689
344 CGCTTAAGTTTCCCAACGCTTCGGCGGTAGCTTCGAACTCGATTC 295
690 Pro.....SerAlaProAlaAlaIleSerSerThrTyrTh 701
294 GTCTGTTGCTTCAACGGCAATCACGACCGACTGAAGTACGCGCTTTTCGC 245
701 rSerSerProSerAlaProValAlaValSerSerThrTyrThrSerSerP 718
244 CGCGATGATGGCGGATTTTCCTTGAACNGCGCGGTAAACACACCAG 195
718 roSerAlaProAlaAlaIle.....SerSerThrTyrThrSerSer 731
194 CGCGATNCTTTTGTCTCAACACGACTTGGCTTTTTCGACGGCATC 145
732 ProSerAlaProValAlaValSerSerThrTyr.....ThrSerSe 745
144 GCCTCTCTTGACTTTTCATCAGGGCGCATACCGGCA..... 108
745 rProSer.....AlaProAlaAlaIleSerSerT 755
107 .....TATTCTTCGCAACGACGACTTCGTA 78
755 hrTyrThrSerSerProSerAlaProValAlaVal 766
```

[illegible]

R. Churchel, C.M.
submitted to the EMBL Data Library, September 1995

A:Reference number: S59302
 A:Accession: S59310
 A:Molecule type: DNA
 A:Residues: 1-1104 <CHU>
 A:Cross-references: EMBL:254141; GSPDB:GN00013; MIPS:YMR317w
 A:Experimental source: strain AB972
 C:Genetics:
 A:Gene: MIPS:YMR317w
 A:Map position: 13R

alignment_scores:
 Quality: 131.00 Length: 466
 Ratio: 0.577 Gaps: 16
 Percent similarity: 48.712 Percent identity: 22.532

alignment_block:

US-09-303-518D-127/rev x S59310

Align seg 1/1 to: S59310 from: 1 to: 1104

```

1331 TCANGGTTTCCACACCTTACGCAACAGCGGCGCCTATTCGTATTGCG 1282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
350 SerSerThrSerSerValSerSerGluAlaProSerSer..... 363
1281 CGGCAGACGAAGCTGCACAAAGCGAGGTCTCTCTCGTCCAAATGCCAAGC 1232
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 .....ThrSerSerValSerSerGluA 372
1231 AACCAATGCTTGGCGGTGTCGATCGCGAGGATTAATCGCGCAA 1182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 laProSerSerThrSerSerValSerSerGluAlaProSerSerThrLys 388
1181 AGCAGGTAGCAGAGTGTACGGGATTTACGGCTCGTAAGTACCAAT 1132
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 SerSerVal.....MetSerSerGluValSerSerAla.....Th 400
1131 CGGCACCATGCGCGGTGCGCCACCGTTCACGCTGCTGTAACCTGAAGA 1082
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 rSerSerValSerSerGluAlaProSerAlaIleSerSerLeuAla 417
1081 GT.....TTGTTTTCAGGAATGGCGAGGTCGTACGGGTGATG... 1041
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
417 erSerArgLeuPheSerSerLysAsnThrSerValThrSerThrLeuVal 433
1040 .....GAGTATTGTCGGGTGCGGCCACCCAGCGACACACTCTTT 997
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
434 AlaThrGluAlaSerSerValThrSerSerLeuArgProSerSerGluTh 450
996 GTCGGCGCTTCTTCGATAACGGAATCTGATTGTTGCTAGCGTCCCAAT 947
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 rLeuAlaSerAsnSerIleIleGlu..... 458
946 AATCGTGGCGCTTGTGTAATCGCGCGTTCATACCGAACCGGAATC 897
458 ..... 458
896 AGCGGTTGTCGTGCTCAACCAATTCGCCCGCAGTAATTCGGATACTTT 847
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
459 SerSerLeuSerThrGlyThrAsnSerThrValSerThrThrThrSerAl 475
846 CGCACCAACCGGTACGCAAGAGCGGTGTTGTTGTTGCTAGGAGAACCC 797
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
475 alaSerSerThrLeuGlySerLys..... 483
796 CCAAGCAATACGCGCTCGGTGTTTCAGACGCGCTGTTGCAACCAACGCT 747
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
484 .....ValSerSerAsnSerArgMetAlaThrSerLys... 495
746 CCGATGCGCAATTACATCTTGATAATGATGTTCCAAACGGTTTGTGTTT... 699
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
496 ...ThrSerSerThrSerSerAspLeuSerLysSerSerValIlePheGl 511

```

```

698 .....GCACGACGCGCTCAATGAATGAATGATGT 671
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
511 yAsnSerSerThrValThrThrProSerAlaSerIleSerLeuThrA 528
670 GGTGCCCACTCAAAACGCGCGGATGCGCGCGGAAATTCATGTGTTTCG 621
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
528 laseProLeuProSerValTrpSerAspIleThrSerSerGluAlaSer 544
620 ATGTTGCGGACGATTTTCAGACGCGACGCTCGCGCAGCTGCTTACACAC 571
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
545 SerIleSerSerAsnLeuAlaSerSerSerAlaProSerAspAsnAsnSe 561
570 ATGGATT.....TTACGCTCGGTCAAAACGGCTCAATACCAACA 533
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561 rThrIleAlaSerAlaSerLeuIleValThrLysThrLysAsnSerValV 578
532 NACNTCGTCTGAAATCNCGCGGCTTCTTGTGATCACAACACACAGGCTCT 483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
578 alSerSerIleValSerSerIleThrSerSerGluThrThrAsnGluSer 594
482 ...GCCGACGCGGATTTGCTCCATCGCATGACGAAAGATGGCGACGG 436
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
595 AsnLeuAlaThrSerSerThrSerLeuLeuSerAsnLysAlaThrAlaA 611
435 CTCGCGATCGACGCGAGGATTTGCTGAACGCGCGGNTACGACCGCGAG 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
611 gSerLeuSerThrSer.....AsnAlaThrSerAlaSerAsnV 624
385 TCCACAAACCGGATTTGATCATGTCNNCNCNNTTCNTCGCGCTTAAG 336
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
624 alProThrGlyThrPheSerSerMetSerSerHisThrSerValIleThr 640
335 .....TTTCCCAACGCTTCGGCGCGTAGCGTTCGAACTCGATT..... 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
641 ProGlyPheSerThrSerSerAlaSerLeuAlaIleAsnSerThrValVa 657
296 ....TCGCTGCTGCTTCAACGCGAATCAGCAGCGACTGAAGTACGCGCT 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
657 lSerSerSerLeuAlaGlyThrSerPheSerThrProGluSerSerProT 674
250 TTTGCGCGGATGGATGGCGGATTTTCCTGAAACGCGCGCTAAAC 201
674 hrThrSerThrLeuValThrSerGluAlaProSerThrValSerSerMet 690
200 ACCACGCGCGATTCCTTTTGTCT.....TCAA 172
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
691 ThrThrSerAlaProPheIleAsnAsnSerThrSerAlaArgProSerPr 707
171 CACACTTGGCTTTTTCAGCGCATCGCTTCC..... 138
707 oSerThrAlaSerPheIleThrGluSerThrSerIleSerSerValP 724
137 .....TTGACTTTTCATCAGGCGCATACCG 111
724 roLeuAlaSerGlyAspValThrSerSerLeuAlaAlaHisAsnLeuThr 740
110 GCATATTCTTCGCAACGCGACTTCGGTAATGACGGCGCGCTCA 63
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
741 ThrPheSerAlaProSerThrSerSerAlaGlnLeuValSerLysSer 756

```

seq_name: pir2:I47141

seq_documentation_block:

gastric mucin (clone PGM-2A) - pig (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C>Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 03-Nov-2000

C:Accession: I47141; S55315

R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.

Gastroenterology 106, 200, 1994

A:Title: Pig gastric mucin: isolation and characterization of a cDNA clone with a nov

A:Reference number: I47141; MUID:94102478

A:Accession: I47141

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-2232 <GSI>
A:Cross-references: EMBL:U08046; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

alignment_scores:
Quality: 124.50 Length: 485
Ratio: 0.571 Gaps: 18
Percent Similarity: 44.948 Percent Identity: 22.062

alignment_block:

US-09-303-518D-127/rev x T34434 ..

Align seg 1/1 to: T34434 from: 1 to: 2232

```
1340 CTTCTCTCTCAANGTTTCCAGCACCTTACGCCAACAGC.....GGGCC 1297
||||| |||.....: |||
670 ProSerGlySerLeuGlyThrGlnSerThrAsnSerProSerPr 686
||||| |||.....: |||
1296 NTATTCCTATTGCGCGGAGAGAGCTGCACAAAGCAGGCTCTTCTT 1247
||| |||.....: |||
586 oSerLeuSerProSerThrSerGlyMetSerThrLeuThrSerGlup 703
||||| |||.....: |||
1246 CGTCAATTCACCAACCAATGTCGCGCTGTCGGTA..... 1206
||| |||.....: |||
703 roSerProSerThrGlnSerSerGlyAlaGlnSerThrLeuThrThr 719
||||| |||.....: |||
1205 ...TCGCGGACGATTAATCCGCAAAAGCAGGCTAGGAGGTCTAG 1159
||||| |||.....: |||
720 ProSerProAsnProSerGlnSerThrSerSerLeuGluSerThrThr 736
||||| |||.....: |||
1158 CGGATTAACGCGCTCGTAAGTACCAATCGGCACCATGGCGGTCGCCAC 1109
||||| |||.....: |||
736 rGlyAlaThrThrSerSerGlySerAlaGlyThrThrMetThrSerPro 753
||||| |||.....: |||
1108 CGTTGACGCGCTGCTGAACCTGAAGAGTTGTTTTCAGAAATGCCG 1059
||||| |||.....: |||
753 erGlnSerSerValGly..... 759
1058 AGGTCGTACGCGTGATGGAGTATTGTCGGCTGCGCGCAACCCAGCC 1009
||||| |||.....: |||
760 .....SerSerGlnGlySerThrSerPr 767
1008 GAACAGCTCTTTGTCGGCGCTTCTTCGATAACCGAAATCTGATGTGGT 959
||||| |||.....: |||
767 oAlaAlaSer..... 770
958 AGCGTCCCAATATCGTGGCGCTTGTGTAATCGCGCGCTTCAATACC.. 909
||||| |||.....: |||
771 .....ThrThr 772
908 GAACCGGAATACACGCGTTGTCGCTCAACCAATTCGCCCGCAGTAAT 859
||||| |||.....: |||
773 SerGlyGluMetThrSer...GlnGlySerThrGlnThrProGlySerSe 788
||||| |||.....: |||
858 TTGCATCTCTTCGACCCCAACCGGTACGCAAGGCGTGGTTGTGA 809
||||| |||.....: |||
788 rValSerThrSerAlaAlaLeuLeuSerThrGlnGlnSerValSer 805
||||| |||.....: |||
808 CTTGAGAACCCCAACCAATACGCGCTCGGTGTTTCAGACGCGCTGTT 759
||||| |||.....: |||
805 hrAsnSerProGlySerThrValThr..... 813
758 GCAACAAACGCTCCGATGGCAATTAACATTTGATAATTTGATGGTCCAAAC 709
||||| |||.....: |||
814 .....ArgProSerThrValSerGlySerThrSerSerGlySerThr 827
708 GGTGTTGTTGACCGCACC.....GCCTCAATGAATGA 674
```

seq_name: p1r2:A40670

seq_documentation_block:

nuclear envelope protein POM 121 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 05-Nov-1999

```

737 .....ThrAlaThrAl 740
780 CTCGGTGTTCAGACGGCGTGTGCAAACAAACGTCGCGATGGCAATTACAT 731
    ::::: :::::
740 aHisThrPheLysProIlePheGluSerValGluProPheAla..... 754
730 CTTGATAATTGATGCTCCAAACGGTTTGTTCACCGCAGCGCTCAATG 681
755 .....AlametProLeuSerProPheSerLeu 764
    ::::: :::::
680 AAATGAATGTCGTGCCACTCAAACCGCGCGATGGCGCGCGCAATTC 631
    ||| :::::
765 LysGlnThrAlaProAlaThrAlaAlaThrSerAlaPro..... 779
    ||| :::::
630 ATGTGTTTCGATGTGGCAGCATTTTCAGACGCAGCTCTGGCGCAGCTG 581
    ::::: :::::
780 .....LeuLeuThrGlyLeuGlyThrAlaThrSerThrValAlat 793
580 CCTTACACACATGGATTTCAGCTCGGTCAAACGGCTCAATACCAGCANA 531
    :::::
793 hr.....GlyThrAla 797

```

[illegible]

```
seq_documentation_block:
```

```
seq_documentation_block:
  glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
  N:Alternate names: extracellular glucamylase; mucin-like protein MUC1; protein YIRO1
  C:Species: Saccharomyces cerevisiae
  C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
```

C:Accession: S48478; A26877; B26877; S27281; JG6123
 R:Rowley, K.
 submitted to the EMBL Data Library, October 1994
 A:Reference number: S48478
 A:Accession: S48478
 A:Molecule type: DNA
 A:Residues: 1-1367 <ROW>
 A:Cross-references: GB:Z47047; EMBL:238061; NID:g603997; PID:g763364; GSPDB:GN00009; MIP
 R:Yamashita, I.; Nakamura, M.; Fukui, S.
 J. Bacteriol. 169, 2142-2149, 1987
 A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
 A:Reference number: A91831; MUID:87194600
 A:Accession: A26877
 A:Molecule type: DNA
 A:Residues: 1-242 <YAM>
 A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
 A:Accession: B26877
 A:Molecule type: DNA
 A:Residues: 762-1331 <YAM>
 A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
 R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
 A:Reference number: S27281; MUID:89031230
 A:Accession: S27281
 A:Molecule type: DNA
 A:Residues: 1-31 <PAR>
 A:Cross-references: EMBL:X13957; NID:g4551; PIDN:CAA32069.1; PID:g4552
 R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A:Title: MucL, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
 A:Reference number: JG6123; MUID:96323237
 A:Accession: JG6123
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1367 <LAM>
 A:Cross-references: GB:U30626; NID:g1304386; PIDN:AAAC49609.1; PID:g1304387
 C:Genetics:
 A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
 A:Cross-references: MIPS:Y1R019C; SGD:S0001458
 A:Map position: 9R
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
 F:5-21/Domain: transmembrane #status predicted <TM1>
 F:1350-1366/Domain: transmembrane #status predicted <TM2>

alignment_scores:
 Quality: 123.00 Length: 532
 Ratio: 0.498 Gaps: 23
 Percent Similarity: 46.429 Percent Identity: 20.489

alignment_block:
 US-09-303-518d-127/rev x S48478 ..
 Align seg 1/1 to: S48478 from: 1 to: 1367

```

1337 TCCTCTCAAGGTTCCAGCAGCTTACGACAGCGGCGCCNTATTCGTA 1288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
891 SerPheSerThrGlyThrValThrValThrProSerSerSerLysThr 905
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1287 TTTCCCGGCGAGCAAGCTGCACAAAGGAGGTCTCTTCGTCCAAAT 1238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
906 ....ProGlySerGlnThrGluThrSerValSerSerThrGluThr 921
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1237 CCAGCAACCAAGTCTGGCGGTGTCGTA.....TCGCG 1200
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
921 hrIleValProThrLysThrThrSerValThrProSerThrThr 937
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1199 ACGATTAAATCGCCAAAGCAGGTAGGAGGATGTCAGCGCATAC 1150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
938 ThrIleThrThrValCysSerThrGlyThrAsnSerAlaGlyGluTh 954

```

```

1149 GCGCTCGTAAGTACCAATCGCAGCAGCGCTGCGCCACCGTTCAGCG 1100
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
954 rThrSerGlyCysSerProLysThrValThrThrValProThrThr 971
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1099 CTGTCGTGAACCTTGAGAGAGTTGTTTTCAGAAATGCGCGAGGTGTA 1050
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
971 hrThrThrSerValThrThr.....SerSerThrThrThrIle 983
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1049 GCGGTGATGAGTATTGTCGCGTGGCGCAACCGCGGAGCGGAGCGTC 1000
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
984 ThrThrThrValCysSerThrThrThrAsnSerAlaGlyGluThrSe 1000
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
999 TTTGCTGCGGCTTCTTCGATACGGAATCTGATTGCTAGCGTCCCA 950
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1000 rGlyCysSerProLysThrIleThr..... 1008
949 AATAATCGTGGCGCCCTGTGTAATCGCGCGTTCATACCGAACCGGAA 900
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1009 ....ThrThrValProCysSerThrSerProSerGluThrAlaSerGlu 1023
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
900 ..... 900
1024 SerThrThrThrSerProThrThrProValThrThrValValSerThr 1040
899 .....ATCACGGGTTGCTGCGTCAACCAATTCGCCGCGAGTAATT 857
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1040 rValValThrThrGluThrSerThrSerThrLysProGlyGlyGluIle 1057
856 GCGTACTTTTCGCCACCCAAACAGGTAGCAAGAGCGGTGTTGTTGACT 807
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1057 hrThrThrPheValThrLysAsnIle...ProThrThrThrLeuThrThr 1072
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
806 TGAGAACCCACCAAGCAATCACGCGTGGTGTTCAGACGGCGCTGTTGC 757
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1073 IleAlaProThrProSerValThr.....ThrValTh 1083
756 AAACAAACGTCGATGGCAATACATCTTGTATTAATGATGTCCAAACGG 707
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1083 rAsnPheThrProThrThrIleThrThr.....ThrV 1094
706 TTTTGTTCACCGCGCGCTCAATGAATGAATGTGCGTCCCACTCAA 657
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1094 alCysSerThrGlyThrAsnSer.....AlaGlyGluThr 1105
656 CGGCGCGGATGCGCGCGCGCAATTCATGTTTCGATGTTGGCAGCAAT 607
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1106 ThrSerGlyCysSerProLysThrValThrThrValProCysSerTh 1122
606 TTCAGACGGC.....ACGTCTGCCAGCTGCCTTACACACATGGA 566
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1122 rGlyThrGlyGluThrThrGluAlaThrThrLeuValThrThrAlav 1139
565 TTTTACGCTCGGTCAACCGCTCAATACCAACANACNTCGTCTGAAATCN 516
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1139 alThrThrValValThrThrGluSerSerThr.....Gly 1151
515 TCGCGGCTTCTTTCATCACACACAGGTCTGCCGCGNAGCGGATGGT 466
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1152 ThrAsnSerAlaGlyLysThrThrGlyThrThrLysSerValPr 1168
465 GTCCATCGCATTCAGAGAGATGCGCAAGCGGTCTGCGCATCGAGCGG 416
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1168 oThrThrThrValThrLeuAlaProSerAlaProValThrPro.... 1183
415 TTTTGTGAACGCGAGGNTACGACGCGAGTCCACAAACCGGATGGATC 366
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1184 .....AlaThrAsnAlaVal.....ProThrThrIle 1192
365 AGATTGCNNNNANTTCGCGCGTAAAGTTGCCAAGCGTTCGGCGC 316
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1193 ThrThrThrGluCysSerAlaAla.....ThrAsnAlaAlaGly.. 1205
315 GTAGCGTTCGAACCTCGATTTCCTCG..... 291

```



```
512 ....NCGGCTCTTTGATCATCAACACACAGGCTCTGCCGACCGGATTCG 467
3158 rProThrAlaThrMetSerThrAlaThrProSerSerThrProGluThrV 3175
466 TG.....TCCATCGCATTGACGAGATGGCGAAGGCTCGGCATCGCAGC 423
3175 aHisThrSerThrValLeuThrThrThrAlaThrThrThrGlyAlaThr 3191
422 GCAGGATTTCGTGAACGAGGNTACGACGCGAGTCCACAAA..... 378
3192 GlySerValAlaThrProSerSerThrProGlyThrAlaHisThrThrLy 3208
377 .....CCGATTGATCAGATTGCNCCNNTTCNTCCGCGCTTAAGTTG 332
3208 sValProThrThrThrThrGlyPheThrAlaThrPro..... 3221
331 CCAACGCTTCGGCGGTAGCGTTCGAACATCGATTTCGTGTCCTTCA 282
3222 .....SerSerSerProGly 3226
281 ACGGCAATCAGACGACTAGTACGCGTTTTCGCCGCGATGGATGC 232
3227 ThrAlaLeuThr.....ProProValTrpIleSe 3236
231 GCGGATTTTCCTGAACGCGCGTAAACACACGCGCGGATNCTTTT 182
3236 rThrThrThrThrProThrThrThrThrProThrThrSerGlySerThrV 3253
181 TGCTCTTCAACACAGACTTGGCCTTTTGTACGCGATCGCCTTCCTTGACT 132
3253 alThrProSerSerIle.....ProGlyThrThr 3262
131 TTCATCAGGCGCATACCGCATATCTTCGCCAAGCAACGCGACTTC 82
3263 His.....ThrAlaArgValLeuThrThrThrThrThrValAlaThrGI 3278
81 GGTAAATGACGGCGCGTCATAAATGACT...TGCTCGGTCCTGCC 39
3278 ySerMetAlaThrProSerSerThrThrGlnThrSerGlyThrPro 3293
seq_name: p1r2:T39903
seq_documentation_block:
serine-rich protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C.Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C.Accession: T39903
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z21889
A:Accession: T39903
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-534 <LYN>
A:Cross-references: EMBL:AL033534; PIDN:CAA22127.1; GSPDB:GN00067; SPDB:SPBC215.13
A:Experimental source: strain 972h-; cosmid c215
C:Genetics:
A:Gene: SPDB:SPBC215.13
A:Map position: 2
C:Superfamily: pig submaxillary mucin
alignment_scores:
Quality: 121.00 Length: 420
Ratio: 0.582 Gaps: 15
Percent Similarity: 49.524 Percent Identity: 19.762
alignment_block:
US-09-303-518d-127/rev x T39903
Align seg 1/1 to: T39903 from: 1 to: 534
1280 GGGCAGCAGAAAGCTGCACAAAGAGGAGTCTTCTTCGTCCTCAATCCAAAGCA 1231
```

```

281 .....SerSerProThrSe 285
336 GTTTCACAGCTTCGGCGCGTAGCTTCGAACTCGATTTCGTCTGC 287
: : : : : : : : : : : : : : : : : : : : : : : : : : :
285 rThrSerSerThrIleSerSer.....SerSerSerSerSerSerP 300
286 CTTCAAGCGCAATCAGACC.....GACTGAAGTACGGGC 252
|| : : : : : : : : : : : : : : : : : : : : : : : : :
300 roThrSerThrSerSerThrIleSerSerSerSerSerSerSer 316
251 TTTTCGCGCGATGATGCGCGCGATTTCCTGAAACNGCGCGTAAA 202
|||| : : : : : : : : : : : : : : : : : : : : : : : :
317 PheSerThrLeuSerSerSerMetSerSerSerSerPheSe 333
201 CACCACGCGCGATNCTTTTCTCTCAACAGCACCTTGGCCTTTTGA 152
: : : : : : : : : : : : : : : : : : : : : : : : : :
333 rSerSerPro.....ThrSerSerSerSerThrIleSerSerS 347
151 CGGCATCGCTTCCTTGACTTCATCNAGGGCGCATACCGGCATATTC 102
: : : : : : : : : : : : : : : : : : : : : : : : : :
347 erSerSerProSerSerPheSerSerThrThrSerSerSerLys 363
101 TCGCCAAGCAAGCGACTTCGATATGACGGCGCGTCAATAGACTTG 52
|| : : : : : : : : : : : : : : : : : : : : : : : :
364 SerSerSerSerPheSerSerThrValSerSerSerSerThrSer 380
51 CTCGCGTCTG 42
|| : : : :
380 rSerThrLeu 383
seq_name: pir2:D83208

```

```

seq_documentation_block:
probable ferredoxin PA3491 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83208
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoquchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: D83208
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <STO>
A:Cross-references: GB:AE004770; GB:AE004091; NID:g9549633; PIDN:AAG06879.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3491

```

```

alignment_scores:
Quality: 121.00 Length: 484
Ratio: 0.540 Gaps: 23
Percent Similarity: 46.281 Percent Identity: 21.901

```

alignment_block:

```

US-09-303-518D-127 x D83208
Align seg 1/1 to: D83208 from: 1 to: 774
28 CTCGCCATCGGGCAGACCGGAGCAGTTCATTTAGACGGCGCGTCAT 77
||||| : : : : :
23 LeuProIleGlnAlaProLeuAlaGlnArgTyrIleValPro..... 37
78 TACCGAAGTCGCGTTCGTCGGAAGATAT...GCGGGTATCGCCCT 124
||||| : : : : :
38 .....LeuGlyGlnHisIleGlyAlaProAlaArgProc 49
||||| : : : : :
125 NGATGAAGTCAAGGAAGCGATGCGCGTCAAAAAGGCCAAGTCTGTTT 174
: : : : : ||| : : : : :

```

```

49 ysValGluVal.....GlyClnAlaValLeuLysGlyGlnThrIleAla 63
175 GAAGACAAAAGNATCCGGCGTGTGTTCACCGCGCCNGTTTCAGGCAA 224
: : : : : : : : : : : : : : : : : : : : :
64 LeuProAspGlyThrValSerAlaAlaLeuHisAlaProThrSerGlyTh 80
225 AATCGCGCGCCATC.....CATCGCGCGCAAAACGCGGTAC 259
: : : : : : : : : : : : : : : : : : : : :
80 rValValAlaIleGlyAlaHisProTyrProHisAlaSerGlyLeuProA 97
260 TTCAGTCGCGTTCGCGTTCGCGTTGAAGGCAACGACGAA..... 297
: : : : : : : : : : : : : : : : : : : : :
97 laProAlaIleValIleAlaSerAspGlyLeuGluArgTyrThrGluLeu 113
298 .....ATCGAGTTCGAACGCTACGCGCCCGCAAGCGTTGGCAAACTT 338
: : : : : : : : : : : : : : : : : : : : :
114 HisProCysProAspPheArgAlaGluSerProLeuAlaLeuGluAr 130
339 AAGCGCGGANGAANTNNGNCAATCTGATCCAAATCCGCGTTGTGGACTG 388
: : : : : : : : : : : : : : : : : : : : :
130 gileArgAlaAlaGlyIleGlyLeuGlyGlyAlaGlyPheProThrA 147
389 CG.....CTCGTANCCGTCCTTCAGCAAAATCCCTGCCGTCGATGCC 432
|| : : : : : : : : : : : : : : : : : : : : :
147 laAlaLysLeuAlaAlaArgProAlaGluLysIle..... 158
433 GAGCGGTTCGCCATCTTCGTCATGCGATGGACACCAATCCG...CTNCG 479
: : : : : : : : : : : : : : : : : : : : :
159 .....HisThrLeuValValAsnGlyAlaGluCysGluProTyrIleSe 173
480 GGCAGACCCCTGTGTGTATCAAGAAGCGCNC.....G 514
||||| : : : : : : : : : : : : : : : : : : : :
173 rAlaAspAspLeuLeuMetArgGluAlaThrGlnValLeuGlyGlyI 190
515 ANGATTTCAGACGANGTTCGTCG.....GTATTGAGCCGCTTG 552
||||| : : : : : : : : : : : : : : : : : : : :
190 leAspIleLeuValGlnIleLeuCysProGluGluValLeuValGlyIle 206
553 ACCGAGCGTAATTCATCTGTGTAGGAGCTGGCGCA.....GA 593
: : : : : : : : : : : : : : : : : : : : :
207 GluAspAspLysProGluAlaIleAlaLeuGlyAlaLeuGlyGlyI 223
594 CGTGCCTCTGAAAATGCTGCCAACATCAACATCAATTCGCGCGC 643
: : : : : : : : : : : : : : : : : : : : :
223 uArgProTyrArgIleValAlaLeuProThrArgTyrProSerGlyGly 240
644 CGCAT.....CCGCC...GGT 657
: : : : : ||||| : : : : :
240 luArgGlnLeuIleGlnLeuLeuThrGlyArgGluValProAlaAspGly 256
658 TTGAGTGGCAGCACATTCATTTTCATTCGCGCGTTCGCGCAACAAAC 707
||| : : : : : : : : : : : : : : : : : : : :
257 LeuProAlaAspIleGlyIleLeuCysGlnAsnValGlyThrLeuAlaI 273
708 CGTTTGGACCATCAATATCAAGATGTAATGCCATCGACGCTTTGTT 757
: : : : : : : : : : : : : : : : : : : : :
273 aVal.....HisAspAlaValValLeuGlyArgProLeuI 285
758 CAACAGCGCGTCTGAACACCGGCGCTGCTTGGTGGTCTCTCAA 807
: : : : : ||||| : : : : :
285 leSer.....ArgIleThrLeuAlaGlyGlyAla 295
808 GTCAACAAACACCGCTCTTCGTCACCGTTTGGTGGCAAGTATCGCA 857
: : : : : : : : : : : : : : : : : : : : :
296 LeuGluArgProMetAsnValGluAlaLeuIleGlyThrProValHisGl 312
858 AATTACTCGC.....GGCAATTGTTGACGCAGACA 889
: : : : : ||| : : : : :
312 uLeuLeuAlaPheAlaGlyLeuAlaGluGlyArgLeu.....G 325
890 ACCGCGTATTTCCGCTTCGCTGATTGAACGGCGGCTTACACAAAGCGCG 939
: : : : : ||| : : : : :
325 luArgValLeuMetGlyGly..... 331

```

99 CGAAGAAATATGCCGGTATCGCCCGCCCTNGATGAAAGTCAAGAGGCGCATG 148
 51 userAsnHisAlaGlyAsnProAlaLysProValValSerProGlyAsp 68
 49 CCGTCAAAAAGCCCAAGTGTGTTTGAAGACAAAAAGNATCCGGCGGTG 198
 68 luValLysThrGlyGlnValIleGlyGluProGluGlyPheIleSerAla 84
 99 GTGTTTACCGCGCCNTTTTCAGGCAAAATCCGCCCAATCCATCCGCGCGA 248
 85 TyrLeuHisSerProValThrGlyArgValLeuGluIle..... 97
 149 AAAGCGCGTACTT.....CAGTCGTCGTGTA 274
 98 LysLysIleLeuHisProIleLeuGlyLysProIleGluAlaIleValI 114
 275 TTGCGGTGAAGGCAACGACGAAATC.....GAGTTC 306
 114 leGluArgThrSerAspGluTprValHisIleGluThrGlyAspPhe 131
 307 GAACGTTACCGGCCCGAAGCGTTGGCAACTTA.....ACGGCGA 347
 131 GluArgMetSerLysGluGluLeuGluIleLysLysAlaGlyI 147
 348 NGAANTNNGNCAATCTGATCCAATCCGTTGTGGACTCGCTGGCGTA 397
 147 e.....ValGlyLeuGlyAlaMetPhe 156
 398 NCCGTCGGTTTCAGCAAAATCCCTGCGTCGATCCGCGAGCGTTGCCCATC 447
 156 roThrHisValLysLeuSerProProGluLysValAspThrLeu 172
 448 TTGCTCAATCCGATGACACCAATCCGCTNGCGCAGACCCCTGTGGTTGT 497
 173 IleValAsnGlyAlaGluCysGluProValLeuThrIleAspHisArgLe 189
 498 GATCAAAAGAACCGNGANGATTTCAGACGANGTGTGCTGATTGAGCC 547
 189 uMetLeuGluAlaGluAspIleLeuGlnGlyIleLeuIleMetMetL 206
 548 GTTTGACCGAGCGTAAATCCATGCTGTGAAGCA...GCTGGCGCAGAC 594
 206 ysvaI.....LeuGlyValGlnLysAlaValValGlyValGlu 218
 595 GTGCGGCTCGAAATGCTGCAACATCGAAACACATGAATTCGGCGGCC 644
 1219 SerAsnLysMetAspAlaTyrHisAsnLeuLysLysValPheLysGly.. 234
 645 GCATCCGCGCGTTTGTAGTGGCAGGCACATTCATTTCATTGAGCGCGTG 694
 2235 TyrProValAspValAlaLeuLeuArgThrLysTyr.....ProGlnG 249
 695 GTGCAACAAA...ACGTTTGTGCACATC..... 720
 249 lYalaGluLysGlnLeuIleTyrAlaIleThrGlyArgMetValProArg 265
 721AATTATCAAGATGT 734
 266 GlyGlyLeuProMetAspValGlyValValGlnAsnValGlyThrCy 282
 735 AATGGCATCGGACGTTTGTTCACACAGCCCTCTGAACACCGAGCGG 784
 282 sValAlaValLysGluAlaValValAspGlyLysProLeuValGluArgG 299
 785 TGATTGCTTGGTGTCTCAAGTCAACAAACACCGCCTCTTGGCTACC 834
 299 lYmetThrValSerGlyAspAlaValLysAsnGlnLysAsnLeuIleVal 315
 835 GTTTGGGTGCGAAGTATCGAAATTAAT.....GCCGGCAATTGGT 878
 316 ArgIleGlyThrProValLysAspValIleAspTyrCysGlyGlyIleAs 332

```

940 CACGATTATTTCGGAGCGCTACCACAATCAGATTTCGGTTTATCGAAGAAGG 989
331 ..... 331
990 CGCAGCAAAAGAGCTGTTTCGGTGGTTCGCGCGCAGCGGACAATACT 1039
      ::::||||::: ||:::
332 .....ProMetMetGlyPheAlaLeuProAsp..... 340:
1040 CCATCAGCGGTACGACCCCTCGCCCATTTCTGAAAAACAACACTCTTCAAG 1089
      ||::: ||::: ||::: ||:::
341 .....LeuSerValProLeuIleLys 347
1090 TTCACGACAGCGCTCAACGGTGGCGACCGCCGATGGTGCCG..... 1131
      :::: ||::: ||::: ||::: ||:::
348 ThrCysAsnCysLeuAlaGlyAspAlaThrGluLeuProGluProVa 364
1132 .....ATTGGTACTTACGAGCGCGTAATGCCGC 1159
      ||| ||::: ||::: ||:::
364 lProAlaMetProCysIleArgCysGlyAspCysAlaGlnValCysPro 381
1160 TAGACATCTCGCTACCCCTGCTTTTCGCGGATTAATCGTGGCGATACC 1209
      :::: ||::: ||::: ||::: ||:::
381 alserLeuLeuProGln...GlnLeuHisPhePheAlaLeuGlyAsp... 395
      :::: ||| ||::: ||::: ||::: ||:::
1210 GACAGCGCGCAGCATTTGGTGTCTTTGGAATTGGACGAGAAGACCTCGC 1259
      :::: ||| ||::: ||::: ||::: ||:::
396 GluHisGluGlnLeuLeuAlaHisAsnLeuPheAspCysIleGluCysGI 412
1260 TTTGTGCGAGTTCTGCTGCGCGGCAATACGAAATANGCCCGCTGTGC 1309
      : ||::: ||::: ||::: ||::: ||::: ||::: ||:::
412 yAlaCysAlaTyrValCysProSerSertle.....ProLeuValG 426
1310 GT 1311
426 in 426
seq_name: pir2:E72398
seq_documentation_block:
  Hypothetical protein TM0244 - Thermotoga maritima (strain MSB8)
  C:Species: Thermotoga maritima
  C:date: 11-Jun-1993 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
  C:Accession: E72398
  C:RefSeq: U00096
  C:Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
  C.M.
  Nature 399, 323-329, 1999
  A:title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
  A:Reference number: A72200; MUID:99287316
  A:Accession: E72398
  A>Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-451 CARN>
  A:Cross-references: GB:AE001708; GB:AE000512; NID:g4980740; PID:4980740
  A:Experimental source: strain MSB8
  C:Genetics:
  A:Gene: TM0244
alignment_scores:
  Quality: 120.00 Length: 456
  Ratio: 0.569 Gaps: 18
  Percent Similarity: 46.272 Percent Identity: 20.175
alignment_block:
  US-09-303-518D-127 x E72398 ..
  Align seg 1/1 to: E72398 from: 1 to: 451
49 GAGCAAGTCATTTATGACGGCCCGCTATTACCGAAGTCGGTTGCTGG 98
      :::: ||| :::: ||::: ||::: ||::: ||:::
35 AspLysProIleGluAlaGlyAlaProLeuProGlnLysValPheValPheLe 51

```


[illegible]

300 SerSerThrPheAlaSerSerThrProIleAlaSerSerSerFF

300 SerSerThrPheAlaSerSerThrProIleAlaSerSerSerFF

```

599 rSerGlnSerProAlaProAsn.....ThrGlySerThrP 612
343 CGCTTAAGTTTGGCAAGCTTCGGCGCTAGGTTTCGAACCTCGATTTCG 294
612 roSerGlnThrSerSerGlnSerProSerProSerMetAsnProSerSer 628
293 TCGTTGCTTCA.....ACGGCAATCACGACCGACTGAAGTAC 256
629 SerThrProThrGlySerSerGlnSerThrIleThrProGluGlySerTh 645
255 GCGCTTTTCGCGGATGGATGGCG.....GCGATTTTCGCTG 218
645 rAlaSerSerProThrGlySerThrGlySerThrPheSerValAlaThrG 662
217 AAACNGCGCGGTAAACACCGCCG...GGATNCTTTTGTCTTCAAC 171
662 luValThrSerGlnSerThrValProSerGlySerSerLeuGlyThrGln 678
170 AGCACT..... 165
679 SerThrAsnSerProSerProSerSerLeuSerProSerThrSerG 695
164 .TGGCCTTTTTCAGCGCATCGCTTCCTTCTGACTTTTCATCAGGGCGCA 116
695 yMetSerThrLeuThrSerGluProSer..... 704
115 TACCGGATATTCCTCCGACAGCAACCGGACTTCGGTAATACGGGCGCG 66
705 ..ProSerSerThrGlnSerSerGlyAlaGlnSerThrLeuThrPro 720
65 TCA 63
721 Ser 721

```

seq_name: pir2:G84348

seq_documentation_block:

hypothetical protein Vng1983h [imported] - Halobacterium sp. NRC-1

C;Species: Halobacterium sp. NRC-1

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: G84348

R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithausner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablonc, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li

A;Title: Genome sequence of Halobacterium species NRC-1.

A;Reference number: A84160; MUID:20504483

A;Accession: G84348

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1494 <STO>

A;Cross-references: GB:AE004437; NID:g10581414; PIDN:AA020155.1; GSPDB:GN00138

C;Genetics:

A;Gene: VNG1983H

alignment_scores:

Quality	117.00	Length	529
Ratio	0.574	Gaps	32
Percent Similarity	38.563	Percent Identity	25.520

alignment_block:

US-09-303-518d-127 x G84348 ..

Align seg 1/1 to: G84348 from: 1 to: 494

66 CGGGCCGCTCATATACGAGTCCGCTTCGTCGCGAAGATATGCCGSTA 115

20 ArgAlaArgArgArgArgSerArgAla..... 28

116 TCGGCCCTNGATGAAAGTCAAGNAGCGATGCGGTCAAAAAGGCCAA 165

```

29 .....GluArgAspProGlyArgArg.....G 36
166 GTGCTGTTTGAAGACAAAAGNATCCGGCGTGTGTTCACCGCGCNGT 215
36 lYAlaValLeuGlyAlaArgThrHisAspArgGlyGlyValArgAla... 51
216 TTCAGGCAAAATCGCGCAT.....CCATCGGGCGGAGAAAGCGCGTAC 259
52 ...ArgArgGlyArgArgHisArgProProProLeuArgThrAlaArgG 67
260 TTCAGTCGTCGTGATTCGCTTGAAGCAA.....CGAC 294
67 yArgThrGlyValAspSerAlaAlaArgAspAlaAlaProValProA 84
295 GAAATCGAGTTCCA..... 308
84 rgThrAlaValArgAspGlyGlyGlnThrArgGluCysProSerArg 100
309 .....ACGCTACGCGCGCGAAG 325
101 GlyProValSerValGlyValThrAlaGluProProValArgArgPr 117
326 CGTTGGCAAACTTAAGCGCGGANGA..... 350
117 oValGlyArgLeuLeuArgValLeuArgGlyAlaArgValAspA 134
350 ..... 350
134 rgAlaGlyAlaValArgGlyArgAlaArgAlaAspThrAlaSerGlyArg 150
351 .....ANTNNGNGCAATCTGATCCCAATCCGGTTT...GTGGACTCGC 391
151 ProProArgArgGlyGluThrArgProHisArgAlaAspValGlyCysG 167
392 T.....GCGTANCCGTCG 405
167 yArgProValCysAspArgValGlyAlaAsnArgProAlaGluProArgH 184
406 TTCAGCAAAATCCCTCCGCTGATCGCGAGCGGTCGCGCATCTT..... 449
184 isArgLeuGlnHisCysArgArgLeuArg.....ArgArgLeuValAla 198
450 .....CGTCAATCGAT...GCACACCAATCC... 473
199 LeuGlnAspAlaLeuGlyProArgAlaGlyAspArgGlyHisGlyAla 215
474 .....GCTNGCGCGCAGACCT 489
215 lLeuProProGluValArgLeuArgValProArgValArgGlyArgProV 232
490 GTGGTTGTGATCAAGAACCGCGANGATTTCAGACGANGTNTGCTGTT 539
232 aAlaValAspLeuArgProArgGlyArgArgAspThrArgProAla... 247
540 ATTGAGCCG.....TTTGACCGAGCGTAAATCCATGTT 574
248 LeuAspProGlyArgThrAlaAlaArgAspArgProGluArgPro..... 262
575 GTAAGGCAGCTGGCGCAGACGTCGCTCTGAAAATGCTGCCAACATCGAA 624
263 .....ArgArgArgPro.....ArgValGlyGlyProArgA 273
625 ACACATGAATTCGGCGCGCGCATCCGCGGTTTGTAGTGGCAGCAGCAT 674
273 rgProLeuValArgArgProLeuArg..... 281
675 TCATTTTCATGTAGCCCGTCGG.....TGCAAAACAAACCGTTT 712
282 .....HisLeuValAspArgIleProArgAspCysArgGlySerArgIle 296
713 G.....GACCATCAATTA...TCAAGATGTAATGCCATCGGA 747

```


675 AATGTGCGTGCACATCAAAACGGCGCGGATGCGGGCCGCCGAATTCATGTG 626
227 227
625 TTTCGATGTTGGCAGCATTTTTCACAGCGCAGCTCGCCAGCTGCTTGA 576
::: ||||| ::: ::::: |||||
228 ..AlaThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThrAlaAla... 242
575 CACACATGATTTTACGCTGCTCAAAACGGCTCAATACCAGCANACNTGG 526
::: ||||| ::::: |||||
243ThrThrAlaAlaTh 248
525 TCTGAATCNTCGNGCGGCTCTTTGATCACAACACAGGCTGTCGCCGNA 476
::: ||||| ::::: |||||
248 rThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThrGlySerProThrS 265
475 GCGGATTGTGTCATCGATTGACGAAGATGGCAACGGCTCGGCATCG 426
||||| ::::: |||||
265 erGlySerThrSerThrGlyAlaSerThrSerProSerAlaSer 281
425 ACGGCA...GGGATTTTCTGAACGGAGGNTAGCAGCGCAGCTCCACAA 379
282 ThrAlaThrSerAlaThrProThrSerThrSerThrSerAlaAla... 296
378 ACGGATTGGATCAGATTGCNNCANNNTTCTCCGCCCTTAAGTTTGCCA 329
297AlaThrThrSerThrProThrProThrSerA 307
328 ACGCTTCGGGCGGTAGCGTTCGAACTCGATTCTCGTTCGCTTCCCTCAACG 279
|||::: ||||| ::::: |||||
307 laaIaThrSerAla.....GluSerThrThrGluAlaProThrSer 320
278 GCAATCACACCGCACTGAAGTACGCGCTTTTCGCCGGATGGATGGCGGC 229
::: ||||| ::::: |||||
321 ThrProThrThrAsp...ThrThrThrProSerGluAlaThrThrAlaTh 336
228 GATTTTCCTCGAACNGCGCGGTAAAC..... 201
::: ||||| ::::: |||||
336 rThrSerProGluSerThrThrValSerAlaSerThrThrSerAlaThrT 353
200ACCAGCGCGGATNCTTTTGTCTTCAACACGACACTTG 162
||||| ::::: |||||
353 hrThrAlaPheThrThrGluSerHisThrSerProAspSerSerThrGly 369
161 CCTTTTGTAGCGATCGCTTCTTGACTTTTCATCAGGGGCGCATACC 112
||||| ::::: |||||
370 SerThrSerThrAlaGluProSerSerThrPheThrLeu.....ThrPr 384
111 GGCAATATCTTCGCCAAGCAACGCCACTTCGGTAAATGACGGCGCGTCAT 62
::: ||||| ::::: |||||
384 oSerThrAlaThrProSer.....ThrAspGlnPheThrGlySerSerA 399
61 AAATGACT.....TGCTCGGTCTGCCCGGATGGCG 30
|||
399 laSerThrGluSerAspSerThrAspSerSerThrValProThrThrGly 415

A; Residues: 1-819 <TIGR>
A; Cross-references: GB:U32841; GB:I42023; NID:q1574529; PIDN:AC23331.1; PID:q1574537
C; Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S]
F: 371-435/Domain: ferredoxin 2[4Fe-4S] homology <FER>

[illegible]


```

seq_name: pir2:S39893

seq_documentation_block:
  rnfC protein - Rhodobacter capsulatus
  C:Species: Rhodobacter capsulatus
  C>Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 17-Mar-2000
  C:Accession: S39893
  R:Schmehl, M.; Jahn, A.; Meyer zu Vilsendorf, A.; Hennecke, S.; Masepohl, B.; Schuppler,
  Mol. Gen. Genet. 241, 602-615, 1993
  A>Title: Identification of a new class of nitrogen fixation genes in Rhodobacter capsula
  A:Reference number: S39892; MUID:94088454
  A:Accession: S39893
  A:Molecule type: DNA
  A:Residues: 1-519 <SCH>
  A:Cross-references: EMBL:X72888; PIDN:G435523; PIDN:CAA51399.1; PID:G435525
  C:Genetics:
  A:Gene: rnfC
  C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S] h
  C:Keywords: iron-sulfur protein
  F:/374-438/Domain: ferredoxin 2[4Fe-4S] homology <FER>

```

|||||:.....
430 sProAlaasnLeu.....ProLeuValGln 438
seq_name: pir2:S55316
seq_documentation_block:
mucin (clone PGM-2B) - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 17-Nov-2000
R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Biochem. J. 308, 89-96, 1995
A:Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.
A:Reference number: S55315; MUID:95275264
A:Accession: S55316
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-317 <TUR>
A:Cross-references: GB:U12768; NID:g915204; PIDN:AAC48525.1; PID:g915207
C:Superfamily: pig submaxillary mucin

alignment_scores:
Quality: 109.00 Length: 417
Ratio: 0.599 Gaps: 18
Percent Similarity: 43.645 Percent Identity: 20.384

alignment_block:
US-09-303-518D-127/rev x S55316 ..

Align seg 1/1 to: S55316 from: 1 to: 317
1283 CCCGGCAGACGAGCTCCACAAAGCGAGGCTTCT..... 1248
||| :|||:||||| ||| ||| |||||
51 ProProGluThrHisGlyAlaThrSerSerThrThrSerValG1 67
1247TCGTCCCAATTCCAAAGCAACCAATGCTTGGCGGTGTCGGTATCGC 1202
|||||:|||||: |||:|||||: ||| |||||:|
67 nProSerSerSerSerAlaProThrThrSerAlaThrSerValGlnP 84
1201 CGACGATTAATCGGGCAAGAGAGGTAGCAGAGTCTAGCGGGATT 1152
|||:|
84 roSer.....SerSerGly... 88
1151 AGCGCTCGTAGTACCAATCGGACCATGGCGGCTGCCACCGCTTGAC 1102
||| ||| |||
89SerAlaProThrTh 93
1101 GGCTCTCGTGAACCTGAAGAGTTGTTTTCAGAAATGGCGGAGGTGG 1052
|:|:|:|:|:|:|:|:|:|
93 rSerAlaThrSerValGlnSer..... 100
1051 TACGGCTGATGAGTATTGTCGGCTCGGGCGCAACCCAGCGCAACAGC 1002
|||:|:|:|:|:|:|:|
101SerSerGlySerAlaProThrThrSerAla 111
1001 TCTTTCTCGCGCCCTCTTCGATAACGGAATCTGATTGTTGAGCGTCC 952
||| :|||:|||||
112 ThrSerValGlnProSer..... 118
951 CAATAATCGTCGGCGCTTGTGTAATCGCGCGCTTCAATACCGAACC 902
||| :|||:|:|:|:|:|:|
119SerSerProProIleSerSerThrIleSerValGlnPro. 132
901 AAATCAGCGGTGTCGTGGTCAACCAATTCGCCCGCAGTAATTTGCGAT 852
|||:|:|:|:|:|:|:|:|:|
133SerSerSerSerAlaProThr..... 140
851 ACTTTTCGCAACCAACCGTACGCAAGAGCGGTGTTGTTGACTTGAGA 802
||| ||| :|||:|:|:|:|
141 ThrSerAlaThrSerValGlnSerSerSerGly..... 152
801 ACCACCAAGCAATCACCGCTCGGTGTTTCAGACGCGCTGTGCAACA 752

152 152
751 AACGTCCGATGCAATTACATCTTTCATTAATTCATGTCCTCAACGGTTTG 702
152 152
701 TTTGTCACGACCGGCTCAATGAATGAATGTCGTGCCACTCAACCGGC 652
|||||:|:|:|:|:|:|:|:|:|
153 SerAlaProThrThrSer.....AlaThrSerValGlnProSe 165
651 CGGATCGCGGCGCGCAATTCATGTTTCGATGTTGTCAGCAATTTTCAG 602
|||:|:|:|:|:|:|:|:|:|
165 rSerSerSerProIleSerSerThrIleSerValGlnProSerS 182
601 ACGGCAGCTCGCGCAGCTGCTTACACACATGATTTAGCTCGCTG 552
|||:|:|:|:|:|:|:|:|:|
182 erSerSerAlaPro..... 187
551 AAACGGCTCAATACACGACANACNTCTGCTGAATCNCNGCGGTCTTT 502
|||:|:|:|:|:|:|:|:|:|
188ThrThrSerAlaThrSerValGlnSerSerSerSe 201
501 GATCACAACACAGGCTCTGCCGCGNAGCGGATTTGTTCCATCGCATTC 452
|||:|:|:|:|:|:|:|:|:|
201 rAlaProThrThrSerAlaThr.....SerValG 211
451 CGAAGATGCGGACGCGCTCGCATCGACGCGAGGATTTTCTGCTGAACG 402
|||:|:|:|:|:|:|:|:|:|
211 InProSerSerGlySerAlaProThr..... 221
401 CGSNTACGACGCGCATCCACAAACCGGATTTGGATCAGATTCGNCN 352
|||:|:|:|:|:|:|:|:|:|
222SerAlaThrSer.....ValGlnSerSerSe 231
351 TTCNTCGCGCTTAAGTTTCCACACGCTTCGGCGGTAGCTGCAACT 302
|||:|:|:|:|:|:|:|:|:|
231 rSerSerProIleSerSerThrIleSerVal.....GlnThrSerS 246
301 CGATTTTCGTGCTTCCCTCAACGCGCAATCACGCGACTGAAAGTACGCG 252
|||:|:|:|:|:|:|:|:|:|
246 erSerSerSerProThrThrSerThr..... 256
251 TTTTCGCGGATGGATGGCGGCTTTTGCCT.....GAAACNGCGCGT 205
|||:|:|:|:|:|:|:|:|:|
257SerValGlnProSerSerSerGlySerAl 266
204 AAACACACGCGCGATNCTTTTGTCTTCAACACAGCACTTGCCTTTT 155
|||:|:|:|:|:|:|:|:|:|
266 aProThrThr.....SerAlaThrSerValGlnProSerS 278
154 TGACGCGATCGCTTCTTGCATTCATCAGGGCGCATACCGGCATAT 105
|||:|:|:|:|:|:|:|:|:|
278 erSerSerProIleSerSerThrIleSerValGlnProSerSer 294
104 TCCTCG.....CCAACGACGCTTCTGTAATGACGGCGCGCT 64
|||:|:|:|:|:|:|:|:|:|
295 SerSerAlaProThrThrSerAlaThrSerValGlnSerSerSe 311
63 A 63
311 r 311

seq_name: pir2:I52257

seq_documentation_block:
episialin - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C:Accession: I52257; I65210
R:Yos, H.L.; De Vries, Y.; Hilken, J.
Biochem. Biophys. Res. Commun. 181, 121-130, 1991
A:Title: The mouse episialin (Muc1) gene and its promoter. Rapid evolution of the rep

seq_documentation_block:
hypothetical protein 3 (gyrB region) - Haloferax sp.

C;Species: Haloferax sp.
C;Date: 11-Feb-1993 #sequence_revision 11-Feb-1993 #text_change 18-Nov-1994
C;Accession: C39135
R;Holmes, M.L.; Dyall-Smith, M.L.
J. Bacteriol. 173, 642-648, 1991
A;Title: Mutations in DNA gyrase result in novobiocin resistance in halophilic archaeobac
A;Reference number: A39135; MUID:91100352
A;Accession: C39135
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-437 <HOL>
A;Cross-references: GB:M38373

```

alignment_scores:      Length: 480
      Quality: 108.00
      Ratio: 0.554
      Gaps: 27
Percent Similarity: 40.625      Percent Identity: 23.958

alignment_block:
US-09-303-518D-127 x C39135      ..

Align seg 1/1 to: C39135 from: 1 to: 437

12 CAATAAAGTCTTAACCTGCCATCGGGCGAGACGGAGCAAGTCATTT 61
   ||| ||||| |||:||||:|
48 GlnAPalaArgIySProProAlaArgAlaArgValProGlnLeu.... 62

62 ATGACGCGCCGCTCAATTACCGAAGTCGCCTTGTGGCGAAGAATATGCC 111
   |||:||||| :||| ||| |||||:|
63 ....ArgGlyArgAspPheAlaLeuArgAlaAspArgArgVal....G 77

112 GGTATCGCGCCCTNGATGAAGTCAAGGAAGCGGATCGCGTCAAAAAGG 161
   :|:|:|:|:|:| :|:|:|:|:| :|||
77 LuHisValProLeu.....ArgGlyArgHisProArgValArgArg 90

162 ...CCAAGTGCCT.....GTTTGAAG 178
   |||:|:| :||| :|||
91 valProGlnArgAspGlnAspGlyValAlaProCAATGATGATGArgHisLeuLeuAr 107

```

208	208
575	GTAAGGACGTGGCGCAGAGCTGCCGTCTGAAATGCTCCCAACATCGAA	624
	: :	
209GlyProArgHisargArglu.....	215
625	ACACATGAATTGGCGCCGCATCCGGCGG.	659
	: : : :	
216ArgHisProThrAlaAspValLeuArgGlyGlu	228
660	GAGTGGCAGCACATTCATTTCAATGAGCGGTCGGTGCAACAACACCG	709
	: : : :	
229	ProGlyHisGlyAspGlyHisLeuGluGlyArgGlyArgProAr	245
710	TTTGACCACATAATCAAGATGTAATGCCATCGGACGTTTGTTCGA	759
	: : : : : : : : : :	
245	pGroGlnGlyArgGluAlaGlyArgGlyAlaHis.....	256
760	ACAGCGCGCTGGAACACCGACGCGTGATTCGTTGGTGGTCTCAAGT	809
	: : : : :	
257ProGlnValArgAlaArgIleYrLeuAlaGlyGluAla	271
810	CAA.....CAAACCGCCCTCT.....CGGTACCGTTT	838
	: : : : :	
272	ArgGlyLeuProGluProArgProLeuGlyValArgThrValHisArgL	288
839	TGGGTGGAAAAGTATCGCAATTACTCGGGCGCAATTGGTTGACGCAGAC	888
	: : : : : : :	
288	yGly.....ArgLeuArgGlyArgValGlyGlnAlaGlyProArGr	303
889	AA.....CCGCGTGATTC.....CGTTCGGT	911
303	InvalProGlyAspPheAlaProGlnGlyGluAspSerGluArgArgGlu	319
912	ATTGAACGGCGCATTACACAAGCGCGCAGATTATTGGTAGCGTACC	961
	: : : : : : :	
320	ThrProProArgProHisSerArgGlyArgAspThrGlyAlaHisH	336
	: :	
962	ACAAFCAGATTCCGTTATCGAAGAAGCGCGCAAGAGCTGTTCCGC	1011
	: :	
336	sArgHisTrpargArgArgArgValArgHisArgGluGlyAlaL	353
1012	TGGGTGGCGCGACCGGACAATACTCCATCAGCGTACGACCTCGG	1061
	: : : : :	
353	eu.....ProAla	355
1062	CCATTTCCTGAAAAACAAACTCTTCAAGTTCAACGACGCGTCAACGGTG	1111
	: :	
356	AlaHisProAsp.....Aspargargargargar	365
1112	GGACGGCGGCATGTGCGCATTTGTTACTTACGAGCGCGTAATGCCGCTA	1161
365	gargargAlaHis.....ProAspAlaAlaL	374
1162	GACATCTCGCTACCTCCTCTTTTTCGGCGATTATTCCTCGCGCATACCGA	1211
	: : : : : :	
374	latYrAlaSerValProAlaHisAlaProAlaHisArgGlyArgLeuArg	390
1212	CAGCGCGAAGCATTTGGTGTGTTGGAATTCGGACGAAGAAGACTCGCTT	1261
	: : : : :	
391	ValArgGlySerThrAla.....AlaValProargProLeuProAr..	404
1262	TGTGACGTCTGCTGCCCGGCAAAATACGAATANGGCC	1301
	: :	
405	GlnHisLeuArgArgAspGlyArgGlyArgGlyAlaGlyPro	417

seq_name: pir2:T45134

```
seq_documentation_block:
  hypothetical protein [imported] - Microbacterium ammoniaphilum (
C|Species: Microbacterium ammoniaphilum
C|Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change
```

C:Accession: T45134
 R:Striebel, H.M.; Seiber, S.; Jarsch, M.; Kessler, C.
 Gene 172, 41-46, 1996
 A:Title: Cloning and characterization of the Mami restriction-modification system from M
 A:Reference number: 222923; MUID: 96257250
 A:Accession: T45134
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-529 <STR>
 A:Cross-references: EMBL:X79027; NID:9984667; PIDN:CRA55649.1; PID:g1679831
 A:Experimental source: ATCC 15354

alignment_scores:
 Quality: 108.00 Length: 473
 Ratio: 0.548 Gaps: 24
 Percent Similarity: 41.649 Percent Identity: 24.101

alignment_block:

US-09-303-518d-127 x T45134 ..

Align seg 1/1 to: T45134 from: 1 to: 529

```

27 CQTGCCCCCGCGCCACCGGACCAAGTCAATTTATGACGGCGCCGTCFA 76
|||||:|||||
128 ProAlaArgArgGlyValLeuArgHisArgHisArgAlaArgG1 144
77 TTACCGAAGTCGCGTGTCTGGCGAAGATATGCCGCTATCGCCCTNG 126
|||||:|||||
144 yGluArgGlyArgGlyProArgGlnGlnValProArgGlnHisPro... 159
|||||:|||||
127 ATGAAGTCAGGAAGGCGATGCCGTCACAAAGGCCAAGTCGTGTTGA 176
|||||:|||||
160 .....ArgGlyArgArgAspArgAlaGlyArgProGlyLeuHisArg 173
|||||:|||||
177 AGACAAAAGNATCC .....GGCGTGGTGTATACCGCGCNG 214
|||||:|||||
174 ArgArgArgAlaArgArgGlnGlnGlyArgProGlnValArgHisG1 190
|||||:|||||
215 TTTGAGCAAAATCCGCGCCATCCATCGCGCGGAAAGCGGTACTTCAG 264
|||||:|||||
190 yAspGlnHisArgAlaAspProArgArgProArgAspProArgAlag 207
|||||:|||||
265 TCGTGTGATGCGGTGAAGCAAGCAAGCAATCGAGTCGACGCTA 314
|||||:|||||
207 lyHisHisProLeuArgGluGlyGlnGlnGlnGluAlaArgProLeu 223
|||||:|||||
315 CGCGCCGGAAGC .....GTTGGCAAACTTAAGCGCGCANG 349
|||||:|||||
224 ArgAlaArgGlyGlnGlnGlyProGlyGlyProAlaGlyArgHisPr 240
|||||:|||||
350 AANTNNGNCAATCTGATCAATCCGTTTGTGGACTGCGCTGCGTANC 399
|||||:|||||
240 oAlaLeuGlyAspGluAsp .....Argp 248
|||||:|||||
400 CGTCCGTTACGAAATCCCTGCGTCGATGCGCGAGCCGTTCCGCACTT 449
|||||:|||||
248 roArgAlaArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 264
|||||:|||||
450 CGTCAATGCGATGACACCAATCCGCTNGCGGAGACCCCTGTGTTGT 497
|||||:|||||
265 ArgLeu .....GlnArgArgGlyArgAlaAspGlyCysAr 277
|||||:|||||
498 ....GATCAAGAGCGCGANGATTTCAGACGANGTNTGCTGATTG 543
|||||:|||||
277 gAlaGlnGlyArgAlaArgAlaArgAlaArgAlaGlyGlyGlyLeu 293
|||||:|||||
544 AGCGTTTGACCGAGCGTAAATCCATGTGTGTAAGCAGCTGCGCAGA 593
|||||:|||||
294 ProArgArgGluAspArg ..... 299
|||||:|||||
594 CGTGCCGCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGC 643
|||||:|||||

```

```

300 .....ValArgAlaHisArg .....P 305
644 CGCATCCGCGCGGTTTGAAGTGGCAGCAGCATTCATTTTCATTTGAGCGGTC 693
|||||:|||||
305 roArgProArgGlyArgGlyGlyAlaAlaArgLeuAspArgAlaGlyG1 321
694 GGTCAACAACAAACCGTTTGACCATCAATATCAAGATGTAATTCGCAT 743
|||||:|||||
321 yValGlyAlaArgProAlaArgProArgArgLeuArgHisProArgProA 338
|||||:|||||
744 CGGACGTTTGTGTTTGAACAGCGCTCTGAACACCGGAGCGGTGATTGCTT 793
|||||:|||||
338 laAspLeuProHisGlyGlyPro ..... 345
794 TGGGTGTTTCTAAGTCAACAACACCGCTCTTGGCTACCGTTTGGGT 843
|||||:|||||
346 .....GlnGlyValAlaArgLeuAspHisProGlnGlyL 357
844 GCGAAAGTATCGCAAAATTTACTGCGGGCGAATTGGTTGACGCGACACACCG 893
|||||:|||||
357 euGlu .....GlyAlaAlaGlySerArgArgHisPro 367
894 CGTGATTTCGCGTTTCGCTATTGAA .....CGGCGGATTACACAAAGCG 937
|||||:|||||
368 HisArgLeuArgAlaArgLeuHisGlnGlyArgGlyAspLeuLeuArg 384
938 CGCACGATTTTGGGACGCTACCAATCAGATTTCCGTTTTCGAAAGAA 987
|||||:|||||
384 g .....ProArgArgAspArgLeuGlyArgArg 394
988 GGCGCGAGCAAGAGCTGTTGCGCTGGTTGCGCCGACCGCGGACCAATA 1037
|||||:|||||
394 lyProArgGlnGlyGln .....GlyAlaHisGlyGlyGlnGly 406
1038 CTCATCACGCGTACGACCCCTCGGCCATTTCCTGAAAAACAACTCTTCA 1087
|||||:|||||
407 LeuArgHisAla ..... 410
1088 AGTTCAGCAGACCGCTCAACGGTGGCGACCG .....C 1119
|||||:|||||
411 .....GlyArgArgArgGlyValProSerGlnLeuAlaValLeuA 425
1120 GCATGTT .....GCCGATTGGTACTTACGA 1145
|||||:|||||
425 rgGlnGlyValCysGlnAlaSerAlaThrLeuAlaLeuTrpMetThrSer 441
1146 GCGCGTAATGCGCGTAGACATCTGCTACCTGCTTTTTCGCGGATTTAA 1195
|||||:|||||
442 GlyArg...LeuLeuArgGluLeuAlaGlyLeuLeuCysArgAspServ 457
1196 TCGTCGCGCATACCGAC .....AGC 1215
|||||:|||||
457 alGluSerHisLeuAspValThrTrpArgAlaAspAlaSerIleSer 473
1216 GCGCAGCATTTGGTTTGGATTGGCAATTGGACGAAGAAGACCTCGCTTTGTG 1265
|||||:|||||
474 ArgArgThrTrpThrValMetAsn...ThrArgLeuArgGlnAlaGlnL 489
1266 CAGCTTCGTTCTCCCGGGCAATACGAATANGCCCGCTGTTCGCTAAGG 1315
|||||:|||||
489 ysSerLeuAlaAlaArgProThrThrSerSerSerIleSerArgSer 505
1316 TGCTGGAAA 1324
|||||:|||||
506 CysTrpGln 508

```

seq_name: pir2:T45462

seq_documentation_block:

membrane glycoprotein [imported] - equine herpesvirus 1

C:Species: equine herpesvirus 1

C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jun-2000

C:Accession: T45462

C:Species: *Xenopus laevis* (African clawed frog)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T14578
R:Shah, S.; Tugendreich, S.; Forbes, D.J.
J. Cell Biol. 141, 31-49, 1998
A:Title: Major binding sites for the nuclear import receptor are the internal nucleoporin-binding sites
A:Reference number: Z18147; MUID:98198465
A:Accession: T14578
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1219 <SHA>
A:Cross-references: EMBL:AF045567; NID:G3047239; PID:G3047240; PIDN:AAAC41273.1

alignment_scores:	
Quality:	108.00
Ratio:	0.448
Percent Similarity:	46.705
Percent Identity:	21.124
Gaps:	28
Length:	516

alignment_block:

US-09-303-518D-127/rev x T14578

Align seg 1/1 to: T14578 from: 1 to: 1219

1322 TCACGACCCTTACGCACAACAGCGGCCNTATTTCGTATTTGGCCGGG..... 1278
|||||::: ::::: ||| ||| |||||
312 SerSerAsnThrLysSerThrValProLeuSerSerThrProGlyLeuGI 328
1278 1278

328 yAspilePheLysLysProAlaGlyMetTrpAspCysAspThrCysLeuV 345
1277 ..CAGACGAAGTCGCACAAACGAGGTCTTCTTCGTCCAATTCCAAGCAA 1230
|||::: ||| ::::: :|||::: |||
345 alclnaSnLSalagluValThrLysCysValAlaCysGluThrProLys 361
1229 CCCAATGCTGC.....CGCGTGCGTATTCGCCGACGATTAAATC 1189
|||||::: ||| ::::: ||| |||||
362 ProGlyThrCysIleLysAlaThrLeuLeuIleProSerThrThrysse 378
1188 GCGCAAAACGAGGTAGCAGCATGCTACGGCGATTACGGCTCGTAAG 1139
|::: ::::: |||||::: ||| ::::: |||
378 rleasnProalathrasnthrLeualaphalasercysSerAlaserI 395
1138 TACCAAATCGCACCATGCGCGGTCCGCACCGTGTACGGCTCGTGAAC 1089
::||| ||| ::::: |||||
395 leProasnGluMetPhelysLysPro..... 404
1088 TTGAAGACTTTGTTTTTCAGGAATGGCGGAGTCTGACGCGTGATGGA 1039
::: ::::: |||::: |||||
405MetGlySerTrp6luCysthrValCysHismetGI 416
1038 GTATTTGTCC.....GGCTCGCGCGCAACCCAGCCGA 1007
::: ::::: ||||| ||| ::::: |||
416 naSnLysThrGluaspasnthrCysValGlyCysAlaGluLysProg 433
1006 AC.....AGCTCTTTGCTG..... 993
433 lyThrValLysSerValprothrAlaalaProSerGlyLeuLeuGlyLeu 449
992CGCGCTTCTCGATAACCGAAATCGATGTGGTA 958
450 LeuHisGlnPheLysLysProThrGly.....SertRp.. 460
957 GCGTCCCCAAATAATCGTGGCGGTCTGTGAATCGCGCGCTTCAATACCG 908
.....: ||||| |||||
461AspCysaspValCysLeulle.....GlansnL 470
907 AACCGAAATCACCGGTGTGTCTCGTCAACCAATTCGCCGCGCGTAATT 858
: ||||| ::::: |||||
470 ysProGluAlaAlays.....CyslleAla 478
857 TCGCATACTTTCGCAACCCAAACGGTAGCAAGAGCGGTGGTTTGTGAC 808

|||||::: ||| ||| :::: :||| |:::
479 CysGluSerAlaLysProGlyThrLysAlaGluProLysGlyThrPheAs 495
807 TTGAAGACCCACCAAGAATCACCGCTCGGTGTTCAGACGGCCTGTG 758
495 p-ThrValIysAsnSerValSerValAlaProLeuSerSerGlyLeu 511
757 CAACAACAGTCCGATGGCAATTACATCTTGATAAATTGATGTGCCAACG 708
512 GlyLeuLeuAspGlnPheLysLysSerAlaGlySer...TrpAspCysAs 527
707 GTTTTGTGTG.....CACCGACCGGCTCAATGAATGAATGTGCCTGCC 664
527 pValCysLeuValGluAsnLysProGluAlaThrLysCysValAlaCysG 544
663 ACTCAAAACCGCGGATCGGGCCCGCAATTCATGTCTTCGATGTTGG 614
544 luThrSerLysProGlyThr..LysAlaGluLeuLysGlyPheGlyThr..S 560
613 CAGCATTTTCACAGCGGCACGCTCGGCCAGCTGCCTTACACACATGGATT 564
560 erThrPheSerSerGlyThrAlaAlaProThr..... 570
563 TTAGCTCGGTCAAACGGCTCAATACCAGCANACTCGTCTGAAATCN... 516
571 PheLysPheGlyValGlnSerSerAspSerThrAlaGluLeuLysSerGl 587
515 ...TCNGCGCTTCTTTGATCACACACAGGGTCTCGCGNACGGGAT 470
587 yAlaSerThrSerGlyPheAlaLysSerIleGlyAspPheLysPheGlyL 604
469 TGGTGTCCATCGCATTCAGAAAGATGGCGAACGGCTCGCGCATCG.... 426
604 euValSerAlaSerThrThrGluGluThrGlyLysLysSerPheThr 620
425ACGCGAGGATTTTCTGTAACGG 403
621 PheGlySerSerThrThrAsnGluValSerAlaGlyPheLysPhe...GL 636
402 ACGNTACGACGCGCAGTCCACAACACCGGATTTGATCAGATTGCNNCNA 353
636 yIleAlaGlySerAlaGlnThrLysProAspThrLeuSer...GlnSerI 652
352 NTTCNTCCGCGCTTAAGTTTGCCCAACGCTTCGGCGCGCTAGCGTTCSAAC 303
652 hrThrSerGlyPheThrPheGlySerVal.....SerAsn 663
302 TCGATTTCGTCGTTGCTTCAACGGCAATCACGACCGACTCAAGTACGCG 253
664 ThrValSerLeuAlaPro.....AlaAlaThrSerSerSerThrGl 678
252 CTTTTCCGCGGATGGATGGCGCGGATTTTCCTTGAACNGCGCGGTAA 203
678 yLeuGln.....ValAlaAlaIleAlaAspSerAsnLeuAlaIat 692
202 ACACC..... 198
692 hrThrAlaAlaLeuLysSerAlaGluLysLysAlaGluAlaProThr 708
197 ..ACGCGCGGATNCTTTGTCTTCAACACGACTTGGCCTTTTTCGAC 151
709 IlethrProPheSerPheGlyLysThrAspGlnAsn.....Ly 721
150 GGCATCCGCTTCCTTGACTTTCATCAGGGCGCATACCGGCATATCTT 101
721 sGluThrAlaSerThrSerPheValPheGlyLys.....L 733
100 CGCCAACAACGCGACTTCGTTAATGACGGCGCGCTCATAAATGACTTGC 51
733 ysAspGluLysThrAspSerAlaProThrGlySerSer.....PheAla 747
50 TCCGGTCTGCCCGGATGGGC...AGGTTTAGACCTTTTTF 12

Percent Similarity: 50.000 Percent Identity: 24.803

alignment_block:

US-09-303-518D-127/rev x A41258

Align seg 1/1 to: A41258 from: 1 to: 725

```
618 GTTGCAGCAGTATTTTCAGACGGCAGCTCTGCGCAGCTGCCTTACACACAT 569
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
124 ValThrSerLysPheThrSerTyrIleCys...HisThr 139
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
568 GGATTTTACGCTCGCTCAACAGGCTCAATACC.....AGCANA 531
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 hrAlaIleSerSerLeuSerGluValGlyThrThrValValSerSer 155
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
530 CNTCTCTGTAATCTCCGCTTCTTTGATC..... 498
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
156 SerAlaIleGluProSerSerAlaSerIleIleSerProValThrSerTh 172
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
497 .....ACAA 494
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
172 rLeuSerSerThrThrSerSerAsnProThrThrThrSerLeuSerSerT 189
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
493 CCACAGGCTCTGCCNAGCGGATGTTGTCATCGCATTCAGCAAGATG 444
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 hrSerThrSerProSerSerThrSerThrSerProSerSerThrSerThr 205
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
443 GCGAACGGCTCGGCATCGACGGCAGGATTTTGCTGAACGGACGGNTAGG 394
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
206 SerSerSerThrSerThrSerSerSerSerSerSerThrSerSerSerSe 222
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
393 CAGCGCAGTCCACAAACCGGATGGATCAGA.....TTGC 359
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
222 rSerThrSerThrSerProSerSerThrSerThrSerSerSerLeuThrS 239
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
358 NNCNNANTCNTCCGCGCTTAAGTTTGCAACGCTTCGGCGCGCTAGCGT 309
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 erThrSerSerSerSerThrSerThrSerGlnSerSerThrSerThrSer 255
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
308 TCGAACTCGATTTCGTCGTTCCCTTCAACGGCAATCACCCGACTGAAG 259
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
256 SerSerSerThrSerThrSerProSerSerThrSerThrSerSerSerSe 272
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
258 TACGGCTTTTCGGCGGATGGATGGCGGATTTGCTGCTGAACACNGCG 209
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
272 rThrSerThrProSer.....SerLysSerThrSerA 284
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
208 CGGTAACACACACCGCGGATNCTTTTGTCTTCAACACGACCTGCGCT 159
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
284 laSerSerThrSerThrSerSerTyr.....SerThrSerThrSerPro 298
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
158 TTTTTCAGCGCA.....TCGCTTCCTTGAC 133
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 SerLeuThrSerSerSerProThrLeuAlaSerThrSerProSerSerTh 315
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 TTTTCATCAGGGGGGCGATA..... 114
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
315 rSerIleSerSerThrPheThrAspSerThrSerSerLeuGlySerSerI 332
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
113 .....CCGGCA 108
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
332 leAlaSerSerThrSerSerValSerLeuTyrSerProSerThrProVal 348
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 TATTCCTTCGCAACGACGCGCTTCGGTAATGACGGGCCCGCATAAAT 58
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
349 TyrSerValProSerThrSerSerAsnValAlaThr...ProSer...Me 363
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
57 GACTTGTCTCC 48
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
363 tThrSerSer 366
```

seq_name: pir2:B46629

seq_documentation_block:

mucin 6, gastric (3-repeat clone) - human (fragment)

N:Alternate names: MUC6

C:Species: Homo sapiens (man)

C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C:Accession: B46629

R:Toribara, N.W.; Robertson, A.M.; Ho, S.B.; Kuo, W.L.; Gum, E.; Hicks, J.W.; Gum Jr., J. Biol. Chem. 268, 5879-5885, 1993

A:Title: Human gastric mucin. Identification of a unique species by expression cloning

A:Reference number: A46629; MUID:93194895

A:Accession: B46629

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-505 <TOR>

A:Cross-references: GB:L07518; MID:g292045; PIDN:AAB61945.1; PID:g292046

A:Experimental source: stomach

A:Note: sequence extracted from NCBI backbone (NCBIN:128397, NCBIP:128399)

C:Genetics:

A:Gene: GDB:MUC6

A:Cross-references: GDB:134734; OMIM:158374

A:Map position: lip15.5-llp15.5

C:Keywords: glycoprotein

alignment_scores:

Quality: 107.00 Length: 310

Ratio: 0.690 Gaps: 14

Percent Similarity: 50.000 Percent Identity: 22.258

alignment_block:

US-09-303-518D-127/rev x B46629

Align seg 1/1 to: B46629 from: 1 to: 505

926 ATCCGCGCTTCAATACCGAAGAAATCACGCGTTCTGTCGCTCAAC 877

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

87 ValAlaProThrSerThr...ThrThrIleThrProAsnProThrSerTh 102

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

876 CAATTCGCCGCGAGTAATTGCGATACITTCGCACCCCAAAACGGTACGCA 827

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

102 rArgThrArgThrProValAlaHisThrAsnSerAlaThrSerSerArgp 119

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

826 AGAGCGGTGGTGTGACTTGAGACCAACCAACCAATCACGCGCTCG 777

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

119 roProProProPheThrThrHisSerProProThrGlySerSer... 133

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

776 GTGTTTCAGAGCGCTCTTGCACAAACACGTCGATGGCAATTACATCTTG 727

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

134ProPheSerSerThrGlyProMetThrAlaThrSer... 145

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

726 ATAATTGATGTTCCAAACGGTTTGTGTCACCGACGCGCTCAATGAAAT 677

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

146PheLysThrThrThrTyrProThrProSerLeu... 157

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

676 GAATGTGCTGCCACTCAAAACCGCGGATGCGGCGCCGCAATTCATGT 627

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

158ProGlnThrThrPro 162

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

626 GTTTCGATGTTGCGACATTTTCAGACGCGCAGCTCTGCGCCAGCTGCTT 577

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

163 LeuThrHisValProProPheSerThrSerLeuValThrProIleThr... 178

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

576 ACACACATGGATTTTACGCTCGTCAACGGCTCAATACACGACANA... 531

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

179 .HisThrValIleThrProThrHisProGlnMetSerThrSerAlaTyrI 195

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

530CNTGCTGTAATFCN 516

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

195 leHisSerThrProThrGlyThrIleAlaSerProThrThrValLysala 211

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

515 TCGNCGGCTTCTTGTATCATCAACACGAGGTCTGCCGACGCGGATGGT 466

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

212 ThrArgSerThrTyrThrAlaProLeuMetThrAlaThrThrSerArgil 228

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

1218 GCAAGCATTTGGTGTGCTTGGAAATTTGGACGAGAAGAC..... 1254
 394 eAlaIleLeuGlyIleAspGluLeuSerGluGluAspArgLeuValVala 411
 1255CTCGTGTGTTCACGCTTCGTCTGC 1278
 411 spARCAAlaArgIstIleGluArgPheLeuSerGluProPheValVala 427

seq name: pir2:T45463

seq_documentation_block:

```
seq_documentation_block;
membrane glycoprotein [imported] - equine herpesvirus 1
```

C: species: equine herpesvirus

C;species: equine herpesvirus 1
C;Date: 31-Jan-2000 #sequence revision 31-Jan-2000 #text change 17-Mar-2000

C; Date: 31-Jan-2000
C: Accession: T45463

C;ACCESSION: 143403
P:Kirisawa, R : Kobayashi, T : Kawakami, Y : Twai, H.

R; AIIISAWA, R.; KODAYASHI, I.
J. Environ Sci 7 79-87 1996

A. Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equine

A; Title: NUCLEOTIDE SEQUENCING
A: Reference number: 722973

A;Reference Number: 422973
Accession: T45463

A;Accession: T45463

A; Status: preliminary;

A;Molecule type: DNA

A;Residues: 1-867 <KIR>

A;Cross-ref

A; Experimental

C;Genetics:

alignment scores:

cores:	Quality:	106.50	Length:	460
--------	----------	--------	---------	-----

Quality: 100
Ratio: 0.48

Percent similarity:	47.609	Percent Identity:	20.652
RaCtO:	0:480	gaps:	12

alignment block:

alignment_block:
IIS-09-303-518D-127/rev x T45463

Align seg 1/1 to: T45463 from: 1 to: 867

1322 TCCAGCACCTTACGCAACAGCGGGGCNTATTTCGTATTTGCCCGGCAGAC 1273
|||||::: ::||| ||| :::
50 SerSerSerProThrThrSerProProThrThrSerSerSerProProTh 66

1272 GAAGCTGCACAAAGCGAGGTCTTTCGTCCAATCCAAGCAACCCAAATG 1223
|::: || |:: |::: |::: |::: |::: |::: |:::
66 rSerThrHisThrSerSerSerProSerSerThrSerThrGlnSerSerSert 83

1222 CTTGGCGGCTGTCCGTATCG...CCGACCATTAATTCGCGCAAAAGCAGG`1176
:::| | | | | | | | | | |
83 hRALAAlamhrSerSerSerAlaProSerThrAlaSerSerThrThrSer qq

1175 GTAGGCAGGATGTCTAGGGCATTACGGCTCGTAAGTACCAATCGGCAC 1126
::: ::||| ||| ::: ::||| ::::
100 TLeProthrSerThrSerThrGluThrThrThrThrThrProThrAlaSe 116

[illegible]

1075 TTTTCAGGAAATGGCCGAGGGTCGTACGGCTGATGGAGTATTGTCCGGC 1026
122 tymbbaalavalmhmbbaalaaacsmbrsaaalaclymhmbbaalaa 149

1025 TGCGGCGCAACCCAGCCGACAGCTCTTTGCTGGGGCTTCTTCGATAAC 976

720 yserThrGluserThrValThrGlySerSerValSerThrValSerGly 736